

# The Telecommunications and Data Acquisition Progress Report 42-102

April–June 1990

E. C. Posner  
Editor

August 15, 1990



National Aeronautics and  
Space Administration

**Jet Propulsion Laboratory**  
California Institute of Technology  
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

## Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio and radar astronomy, it reports on activities of the Deep Space Network (DSN) and its associated Ground Communications Facility (GCF) in planning, in supporting research and technology, in implementation, and in operations. Also included is TDA-funded activity at JPL on data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Operations (OSO). The TDA Office also performs work funded by two other NASA program offices through and with the cooperation of the Office of Space Operations. These are the Orbital Debris Radar Program (with the Office of Space Station) and 21st Century Communication Studies (with the Office of Exploration).

In the search for extraterrestrial intelligence (SETI), the *TDA Progress Report* reports on implementation and operations for searching the microwave spectrum. In solar system radar, it reports on the uses of the Goldstone Solar System Radar for scientific exploration of the planets, their rings and satellites, asteroids, and comets. In radio astronomy, the areas of support include spectroscopy, very long baseline interferometry, and astrometry. These three programs are performed for NASA's Office of Space Science and Applications (OSSA), with support by the Office of Space Operations for the station support time.

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund which involve the TDA Office are included.

This and each succeeding issue of the *TDA Progress Report* will present material in some, but not necessarily all, of the following categories:

### OSO Tasks:

- DSN Advanced Systems
  - Tracking and Ground-Based Navigation
  - Communications, Spacecraft-Ground
  - Station Control and System Technology
  - Network Data Processing and Productivity
- DSN Systems Implementation
  - Capabilities for Existing Projects
  - Capabilities for New Projects
  - New Initiatives
  - Network Upgrade and Sustaining
- DSN Operations
  - Network Operations and Operations Support
  - Mission Interface and Support
  - TDA Program Management and Analysis
- Communications Implementation and Operations
- Data and Information Systems
- Flight-Ground Advanced Engineering

### OSO Cooperative Tasks:

- Orbital Debris Radar Program
- 21st Century Communication Studies

**OSSA Tasks:**

Search for Extraterrestrial Intelligence  
Goldstone Solar System Radar  
Radio Astronomy

**Discretionary Funded Tasks**

# Contents

## OSO TASKS DSN Advanced Systems TRACKING AND GROUND-BASED NAVIGATION

<b>The First Geocenter Estimation Results Using GPS Measurements .....</b>	<b>1</b>
R. P. Malla and S. C. Wu NASA Code 310-10-61-84-02	
<b>Computer Simulations of Ions in Radio-Frequency Traps .....</b>	<b>10</b>
A. Williams, J. D. Prestage, L. Maleki, J. Djomehri, and E. Harabetian NASA Code 310-10-62-15-00	
<b>Design of a Fiber-Optic Transmitter for Microwave Analog Transmission With High Phase Stability .....</b>	<b>27</b>
R. T. Logan, Jr., G. F. Lutes, L. E. Primas, and L. Maleki NASA Code 310-10-62-16-00	
<b>Navigational Utility of High-Precision Radio Interferometry for Galileo's Approach to Jupiter .....</b>	<b>34</b>
W. M. Folkner NASA Code 310-10-63-88-01	

## COMMUNICATIONS, SPACECRAFT-GROUND

<b>Linear-Phase Approximation in the Triangular Facet Near-Field Physical Optics Computer Program .....</b>	<b>47</b>
W. A. Imbriale and R. E. Hodges NASA Code 310-20-65-04-00	
<b>Site Comparison for Optical Visibility Statistics in Southern Arizona .....</b>	<b>57</b>
K. A. Cowles NASA Code 310-20-67-88-03	
<b>The Theoretical Limits of Source and Channel Coding .....</b>	<b>62</b>
S. J. Dolinar and F. Pollara NASA Code 310-30-71-83-02	
<b>A High-Speed Distortionless Predictive Image-Compression Scheme .....</b>	<b>73</b>
K.-M. Cheung, P. Smyth, and H. Wang NASA Code 310-30-71-83-02	

## DSN Systems Implementation CAPABILITIES FOR EXISTING PROJECTS

<b>Interagency Telemetry Arraying for Voyager-Neptune Encounter .....</b>	<b>91</b>
D. W. Brown, W. D. Brundage, J. S. Ulvestad, S. S. Kent, and K. P. Bartos NASA Code 314-40-41-81-13	
<b>The Parkes Front-End Controller and Noise-Adding Radiometer .....</b>	<b>119</b>
T. J. Brunzie NASA Code 314-30-66-11-11	

## CAPABILITIES FOR NEW PROJECTS

<b>The Impact of NRZ Data Asymmetry on the Performance of a Space Telemetry System .....</b>	<b>139</b>
T. M. Nguyen NASA Code 314-40-22-70-08	

# **OSSA TASKS**

## **SEARCH FOR EXTRATERRESTRIAL INTELLIGENCE**

**Technical Considerations on Using the Large Nançay Radio Telescope for SETI** ..... 152 *25*  
S. Gulkis, F. Biraud, J. Heidmann, and J. Tarter  
NASA Code 314-30-69-10-03

**Referees** ..... 161

27-43  
200-12  
p-9 August 15, 1990

N91-11974

# The First Geocenter Estimation Results Using GPS Measurements

R. P. Malla<sup>1</sup>  
S. C. Wu

Tracking Systems and Applications Section

*The center of mass of the Earth is the natural and unambiguous origin of a geocentric satellite dynamical system. A geocentric reference frame assumes that the origin of its coordinate axes is at the geocenter, in which all relevant observations and results can be referred and in which geodynamic theories or models for the dynamic behavior of Earth can be formulated. In practice, however, a kinematically obtained terrestrial reference frame may assume an origin other than the geocenter. A fast and accurate method of determining origin offset from the geocenter is highly desirable. Global Positioning System (GPS) measurements, because of their abundance and broad distribution, provide a powerful tool to obtain this origin offset in a short period of time. Two effective strategies have been devised. Data from the first Central and South America (Casa Uno) global GPS experiment have been studied to demonstrate the ability of recovering the geocenter location with present-day GPS satellites and receivers.*

## I. Introduction

Reference frames are established in order to represent positions and motions of objects with respect either to the Earth (terrestrial frames) or to celestial bodies in space (celestial inertial frames). One of the geophysical requirements of a reference frame is that other geophysical measurements can be related to it; for example, the reference frame used for expressing the Earth's gravity field as a spherical harmonic expansion adopts the center of mass of the Earth as origin. This frame must be related to the adopted terrestrial frame as well as to the inertial frame in

which the satellite orbits are calculated. The geocenter is at a focus of the orbital ellipse of a geocentric satellite dynamical system. It is therefore directly accessible through dynamical methods. But in practice the origin assumed by a kinematically obtained terrestrial reference frame can be at some location other than at the Earth's center of mass. Any time-dependent offset in a geocentric terrestrial frame origin from the geocenter can be misinterpreted as plate motions. In order to avoid such confusion, it is important to determine as accurately as possible the translational offset of the adopted terrestrial reference frame origin from the geocenter.

Two effective strategies were devised to determine the reference frame origin offset from the geocenter using

<sup>1</sup>Member of Professional Staff, Sterling Software, Pasadena, California.

Global Positioning System (GPS) observations. The results of covariance analyses performed to investigate the accuracy with which the geocenter position can be determined with these two strategies have been published in [1]. The analyses indicate that the geocenter position can be determined to an accuracy of a few centimeters with just one day of precise GPS pseudorange and carrier phase data collected by a global GPS tracking network.

Several regional GPS experiments have been carried out since 1985, but the tracking ground networks have not been extensive enough to provide the global coverage needed for accurate geocenter estimation. The Casa Uno experiment, however, used a semiglobal network stretching over the South Pacific and across the continental United States and Europe, thus providing an opportunity to demonstrate the capability of GPS data for geocenter determination. The analyses of the results using this set of data show that the accuracy is strongly limited by a nonglobal GPS constellation, the received data quality, and the geometrical distribution of the semiglobal ground tracking sites.

## II. Two Effective Strategies to Determine the Geocenter

The two strategies devised in [1] to determine the origin offset from the geocenter are briefly reviewed here. A fiducial network consists of two or more GPS tracking stations whose positions have been determined in an Earth-fixed coordinate frame to a very high accuracy, usually by Very Long Baseline Interferometry (VLBI) or Satellite Laser Ranging (SLR). Several GPS receivers at other, less accurately known stations also observe the GPS satellites along with the fiducial network. The data can be processed simultaneously to adjust GPS satellite states and the positions of the nonfiducial sites. The fiducial stations established by VLBI provide a self-consistent, Earth-fixed coordinate frame; thus the improved GPS satellite orbits and the nonfiducial stations can be expressed with respect to this coordinate frame to a greater accuracy. The filter process is designed so that the baselines between a reference site and all other nonfiducial sites are adjusted along with GPS orbits and the absolute coordinates of the reference site. The correction to the reference site coordinates infers the adjustment of the geocenter position coordinates. This concept has been used in Strategy 1, where one or two fiducial baselines are fixed or constrained by their a priori uncertainties [1]. The orientation of the adopted coordinate frame is defined by the fixed baselines, and the absolute scaling can be fixed either by the length of these baselines or by the Earth's gravitational constant (GM). Both are known to an accuracy of about one part

in  $10^8$ . The absolute scale derived from the fixed baseline length allows the coordinate frame thus established to be consistent with the VLBI frame of the fiducial baselines.

In Strategy 2, only the longitude of a reference site is held fixed; all other site coordinates are adjusted simultaneously along with GPS satellite states. The absolute scale is provided by the Earth's GM. The geocentric radius at a station can be obtained from the adjusted periods of GPS orbits and pseudorange measurements. The time signature of the measurements defines the latitude. The coordinate system thus defined will be an Earth-centered, Earth-fixed (ECEF) coordinate frame.

The covariance study carried out in [1] assumed a full constellation of eighteen GPS satellites distributed in six orbital planes. A data arc spanning over 34 hours from a network of six globally distributed tracking stations was also assumed. For Strategy 1, the a priori uncertainty for the relative positions of the fiducial sites was assumed to be 3 cm. P-code pseudorange and carrier phase data noise were assumed to be 5 cm and 0.5 cm respectively when integrated over 30 minutes and corrected for ionospheric effects by dual-frequency combination. Carrier phase biases were adjusted with a large a priori uncertainty. The abundance and broad distribution of the GPS measurements allow the GPS and station clocks to be treated as white-noise processes and adjusted to remove their effects on the solutions. Also adjusted were the zenith tropospheric delays at all ground sites, which were treated as random-walk parameters to model the temporal variation. Such models have proven to be effective in removing tropospheric errors without heavily depleting the data strength [2]. The same network of six tracking sites was also used to assess Strategy 2. The estimated quantities were the GPS satellite states, the coordinates of all six sites except for the fixed longitude of the reference site, white-noise clocks, random-walk troposphere parameters, and carrier phase biases.

Data arcs of various lengths were used in the covariance analysis. In Strategy 1, at the end of 34 hours the formal error in origin offset from geocenter is 4.0 cm (rms of all three components). The dominating error is due to the assumed error in the fiducial baselines. Any improvement in the baseline estimates will therefore directly benefit the geocenter determination. As the arc length of the data is increased, the error due to data noise is reduced. On the other hand, the systematic error due to the fiducial baseline errors persists. The situation in Strategy 2, however, is different. Here, under the assumptions of the study, data noise is the primary error source, which can be reduced by increasing the data arc length. In reality there

will be other errors, such as multipath, troposphere mis-modeling, and higher order ionospheric effects. The origin offset error from the geocenter—as predicted by the covariance analysis using this strategy—was 2.1 cm at the end of the 34 hours [1].

### III. Methodology for Processing Casa Uno Data

The first Central and South America (Casa Uno) GPS experiment was carried out from January 18 to February 5, 1988. This experiment was the first civilian effort at implementing an extended international (15 nations) GPS satellite tracking network [3]. Twelve globally distributed sites were selected (Fig. 1) to provide improved global coverage for the geocenter study.

The collected data are from seven GPS satellites in the constellation, distributed in two orbital planes with a separation of approximately 120 deg in the right ascension of the ascending nodes. Both strategies are applied to determine the geocenter location. The analyses present the achievable accuracy based upon the available suboptimal tracking conditions during the Casa Uno experiment. For example, the baselines known to higher accuracies are concentrated only within the continental United States and are relatively short compared to the extensive area covered by the tracking network; the seven satellites distributed in only two orbital planes do not provide a good global coverage; and the pseudorange data are not of high quality due to the antennas and receivers used in this experiment.

For Strategy 1 the Owens Valley Radio Observatory (OVRO)-Haystack baseline is held fixed. The other ground station locations, the geocenter location, and the GPS satellite states are adjusted with respect to the baseline reference point OVRO. Thus the coordinate system has the scale and orientation as defined by this fixed baseline, and the adjustment to the geocenter location gives the offset of the coordinate frame origin from the geocenter. Table 1 gives the a priori uncertainties adopted for Strategy 1. In Strategy 2, the longitude of the reference site OVRO is held fixed, and all the other site coordinates along with the GPS satellite states are adjusted simultaneously. Table 2 lists the a priori uncertainties adopted in Strategy 2, which differ from those in Strategy 1 (Table 1).

The coordinate systems defined by Strategy 1 and Strategy 2 have fundamental differences. In Strategy 1 the geocenter offset is determined directly while fixing one baseline; in other words, the estimated correction to the geocenter location represents the coordinate frame offset implied by the two ends of the fixed baseline. In Strat-

egy 2, the coordinate offset is not directly determined. Instead, the corrections to all the tracking site coordinates, except for the reference site longitude, are estimated. The geocenter offset is inferred from these estimates through a constrained seven parameter coordinate transformation. The seven parameters are solved by treating filter estimates of corrections to tracking site coordinates as the measurements, whereas the measurement covariance matrix is represented by the corresponding filter covariance plus the a priori covariance of tracking sites. The arbitrarily fixed reference longitude introduces a small rotation  $R_z$  about the Z axis; this rotation is taken into account by solving for  $R_z$  while the scale factor and the rotations about the X and Y axes are kept fixed. The geocenter offset is represented by three translational parameters which are assigned a priori uncertainty of 10 m. The Appendix gives a full account of the method of seven parameter coordinate transformation.

Since the same tracking network was used for both strategies, they have similar a priori conditions. For example, in both strategies, the a priori geocentric reference frame, in which the nominal station coordinates are expressed, is derived from Goddard global VLBI coordinates for those same sites and is rotated and translated using the results from SLR data for geocentricity.

### IV. Results

Although the Casa Uno ground network was designed to collect a maximum amount of data with the available seven GPS satellites, it provides only a semiglobal coverage. In order to compensate for this suboptimal circumstance, the data strength was increased by using a 5-day data arc, and the geometrical strength of the network was improved by constraining the well-known sites with appropriate a priori uncertainties. The a priori uncertainties assumed for Strategy 1 in this experiment are listed in Table 1. In the covariance analysis presented in [1] the GPS Block I ROCK 4 model for solar radiation pressure was used, where the three parameters  $G_x$ ,  $G_y$ , and  $G_z$  were considered with 10-percent error. The constant acceleration in the y axis  $G_y$  is called the y-bias parameter [4]. In the Casa Uno multiday analysis a new approach was adopted in which two constant solar pressure parameters  $G_y$  and  $G_{xz}$  were estimated along with two tightly constrained process noise parameters  $G_x$  and  $G_z$  [2]. The parameter  $G_{xz}$  is a combined effect of ( $G_x + G_z$ ).

Figure 2 shows the formal rms error in the geocenter solution using Strategy 1. The error associated with the geocenter offset estimation due to data noise is 31 cm and

that due to baseline uncertainty is 7.1 cm, resulting in a total formal rms error in the geocenter offset solution of 31.8 cm. The effect due to data noise dominates the error, showing the poor quality of data; however, the error due to the fixed baseline is also significant. These formal errors are significantly higher than the 4 cm predicted by the earlier covariance study because the ground tracking sites were not uniformly distributed, the fixed baseline was not long enough to provide a global control, and only one-third of the GPS constellation was in place during the experiment. The origin offsets estimated with the Casa Uno GPS data, with respect to the SLR derived geocenter, are  $-142$ ,  $-33$ , and  $-43$  cm along the X, Y, and Z axes (Fig. 3).

Table 2 lists the variation in the a priori uncertainties from Table 1 as applied to Strategy 2. The longitude at OVRO was fixed and all the other stations had large a priori sigmas. The same tracking station network was used in both cases. The estimated parameters include all the ground station coordinates and the GPS satellite states. The formal rms error in the geocenter using this strategy is found to be 21.5 cm, which is due only to the data noise. The estimated origin offsets from the nominal geocenter are  $-96$ ,  $-97$ , and  $-51$  cm along the X, Y, and Z axes (Fig. 3).

The origin solutions from the two strategies differ from each other along X, Y, and Z axes by 46, 65, and 8 cm respectively. The rms of these differences is 45.7 cm, which is the mean origin offset in each component between results from Strategy 1 and Strategy 2. The geocenter estimates from Strategy 1 and Strategy 2 agree at the 2-sigma level. The apparent disagreement in X and Y components between the GPS and SLR solution is puzzling. It may be that the baselines between European and U.S. sites are known less accurately than the 10 cm assumed here. A

recent independent (but inconclusive) experiment gave a baseline adjustment of up to 50 cm between the station at Wettzell and the U.S. sites.

The anticipated improvement in data quality, better GPS constellation, and even distribution of the tracking stations should in the future improve the accuracy of geocenter estimation to the few centimeters predicted by the earlier covariance studies.

## V. Summary and Conclusion

The Casa Uno 5-day GPS data arc, despite the uneven data quality, limited global coverage, and partial satellite constellation, can recover the geocenter position to an estimated accuracy of better than 35 cm using either of two strategies. The two coordinate frames differ from one another by three translational parameters and one rotational parameter about the Z axis. The transformation parameters between the two coordinate frames are very sensitive to the a priori values and constraints applied to the participating sites. The discrepancy between the results from Strategy 1 and Strategy 2 falls within their 2-sigma error.

With future superior receiver data quality, an evenly distributed global network including longer fiducial baselines, and an increased number of satellites distributed in more orbital planes, the results are expected to improve to an accuracy of a few centimeters as indicated by previous covariance studies. Inclusion of the Deep Space Network (DSN) sites will provide the well determined global baselines that are lacking in the Casa Uno network. Longer baseline ties to the well-known sites in North America from Europe and the southern hemisphere will provide a much stronger network geometry for geocenter determination.

## References

- [1] R. P. Malla and S. C. Wu, "Deriving a Unique Reference Frame for GPS Measurements," *TDA Progress Report 42-95*, vol. July-September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 1-13, September 15, 1988.
- [2] W. Bertiger, S. M. Lichten, and E. C. Katsigris, "A Demonstration of Submeter GPS Orbit Determination and High Precision User Positioning," *Proc. IEEE PLANS '88 Position Location and Navigation Symposium*, Orlando, Florida, pp. 185-194, November 1988.
- [3] R. E. Neilan, T. H. Dixon, T. K. Meehan, W. G. Melbourne, J. H. Scheid, J. N. Kellogg, and J. L. Stowell, "Operational Aspects of Casa Uno '88-The First Large Scale International GPS Geodetic Network," *IEEE Trans. of Instr. and Meas.*, vol. 38, no. 2, pp. 648-651, April 1988.
- [4] H. F. Fliegel, W. A. Feess, W. C. Layton, and N. W. Rhodus, "The GPS Radiation Force Model," *Proc. First Int. Symp. on Precise Positioning with GPS*, Rockville, Maryland, pp. 113-120, April 1985.

**Table 1. The a priori assumptions for geocenter study using Casa Uno data (one baseline fixed)**

Reference site:	OVRO
Other fiducial sites:	Haystack
Nonfiducial sites:	Blackbirch, Canberra, Kokee, Samoa, Cocos, Albrook, Mojave, Ft. Davis, Wettzell, and Onsala
GPS constellation:	GPS 11, GPS 3, GPS 4, GPS 6, GPS 8, GPS 10, and GPS 9 distributed in 2 orbital planes
Cutoff elevation:	15 deg
Data span:	up to 5 days
Data interval:	6 min
Data noise:	175 cm pseudorange; 1 cm carrier phase
GPS epoch state:	20 km and 20 m/sec (adjusted)
Geocenter position: (each comp.)	10 m (adjusted)
Baseline coordinates: (relative to OVRO, each comp.)	Haystack 4 cm (fixed); Ft. Davis and Mojave—4 cm (adjusted); Wettzell and Onsala—10 cm (adjusted); Others—1 km each comp. (adjusted)
Solar pressure:	$G_y$ and $G_{xz}$ (adjusted)

**Table 2. Variation of assumptions from Table 1 for Strategy 2 (one longitude fixed)**

Reference site:	OVRO
Reference site coordinates:	10 m (latitude) 0 m (longitude) 10 m (height)
Other site coordinates:	10 m each component

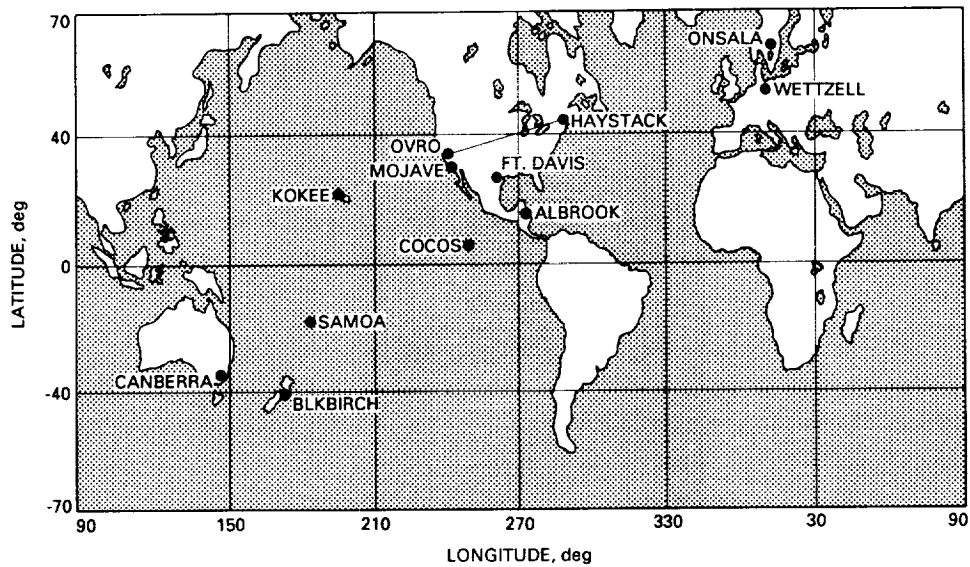


Fig. 1. The Casa Uno tracking sites for geocenter studies.

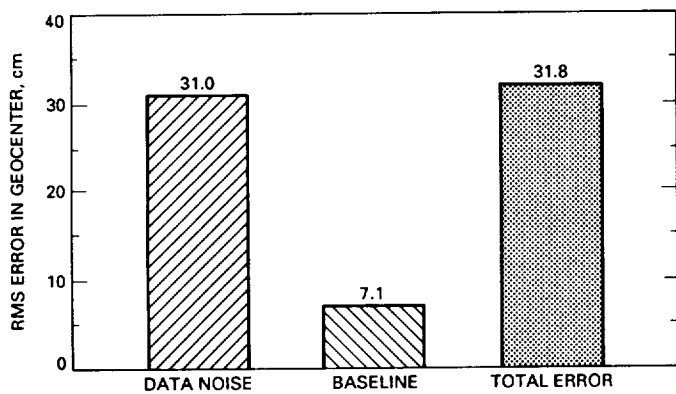


Fig. 2. Geocenter error ( $1\sigma$ ) from covariance study for Casa Uno 5-day arc data using Strategy 1.

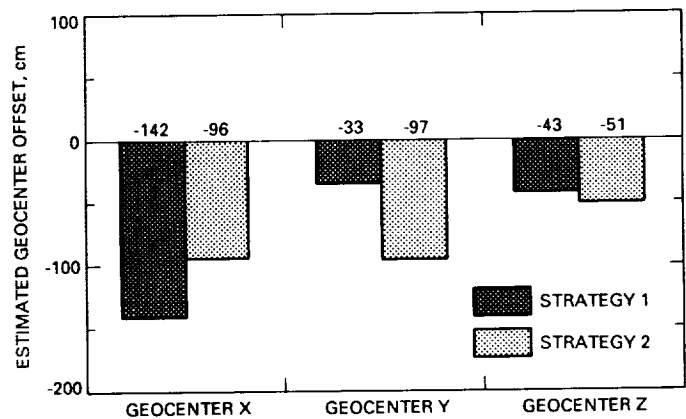


Fig. 3. Geocenter estimates from Casa Uno data using Strategies 1 and 2.

ORIGINAL PAGE IS  
OF POOR QUALITY

## Appendix

### Seven Parameter Coordinate Transformation

The coordinates of a point in a given Cartesian reference frame can be expressed with respect to any other Cartesian reference frame as

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{Ref1}} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + (1 + \Delta L) \mathbf{R} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{Ref2}} \quad (\text{A-1})$$

where  $\Delta X$ ,  $\Delta Y$ , and  $\Delta Z$  are the three translational parameters, and  $\Delta L$  is the scale difference between reference frames Ref1 and Ref2. The rotation matrix  $\mathbf{R}$  represents the rotation required about the three axes in order to align the two reference frames

$$\mathbf{R} = \mathbf{R}(R_y)\mathbf{R}(R_x)\mathbf{R}(R_z)$$

where  $\mathbf{R}(R_i)$  is the matrix representing a right-handed rotation about the  $i$ th axis through an angle  $R_i$ . The rotation matrix  $\mathbf{R}$  is obtained by multiplying the three matrices in sequence, i.e.,

$$\mathbf{R} = \begin{bmatrix} \cos R_z \cos R_y - \sin R_z \sin R_x \sin R_y & \sin R_z \cos R_y + \cos R_z \sin R_x \sin R_y & -\cos R_x \sin R_y \\ -\sin R_z \cos R_x & \cos R_z \cos R_x & \sin R_x \\ \cos R_z \sin R_y - \sin R_z \sin R_x \cos R_y & \sin R_z \sin R_y - \cos R_z \sin R_x \cos R_y & \cos R_x \cos R_y \end{bmatrix} \quad (\text{A-2})$$

For small angle rotations and neglecting products of small angles:

$$\mathbf{R} = \begin{bmatrix} 1 & R_z & -R_y \\ -R_z & 1 & R_x \\ R_y & -R_x & 1 \end{bmatrix} \quad (\text{A-3})$$

Substituting Eq. (A-3) into (A-1) and rearranging the terms gives:

$$\begin{aligned} \Delta X - R_y Z_{\text{Ref2}} + R_z Y_{\text{Ref2}} + (X_{\text{Ref2}} + R_z Y_{\text{Ref2}} - R_y Z_{\text{Ref2}}) \Delta L + (X_{\text{Ref2}} - X_{\text{Ref1}}) &= 0 \\ \Delta Y + R_x Z_{\text{Ref2}} - R_z X_{\text{Ref2}} + (Y_{\text{Ref2}} - R_z X_{\text{Ref2}} - R_x Z_{\text{Ref2}}) \Delta L + (Y_{\text{Ref2}} - Y_{\text{Ref1}}) &= 0 \\ \Delta Z - R_x Y_{\text{Ref2}} + R_y X_{\text{Ref2}} + (Z_{\text{Ref2}} + R_y X_{\text{Ref2}} - R_x Y_{\text{Ref2}}) \Delta L + (Z_{\text{Ref2}} - Z_{\text{Ref1}}) &= 0 \end{aligned} \quad (\text{A-4})$$

This set of equations represents the relationship between two closely oriented, closely scaled orthonormal Cartesian reference frames. Every point appearing in both frames will generate three observation equations, where the seven transformation parameters ( $\Delta X$ ,  $\Delta Y$ ,  $\Delta Z$ ,  $R_x$ ,  $R_y$ ,  $R_z$ ,  $\Delta L$ ) relating the two frames are to be uniquely estimated. A unique solution requires two stations with their coordinates known in both reference frames and only one component of a third station known in both frames. In practice there will be more than three stations participating and the weighted least squares solution gives the solution for the parameters.

Let  $L_b$  represent the observations with corresponding variance-covariance matrix  $\Sigma_L$ , so that the weight matrix  $\mathbf{P} = (\Sigma_L)^{-1}$ . Each observed quantity is expressed as a function of the parameters in the model

$$L_a = F(X_a) \quad (\text{A-5})$$

Here  $L_a$  represents the theoretical values of the observed quantities and  $X_a$  the theoretical values of the parameters. The Taylor series linearization gives the observation equation as

$$V = Ax + (L_o - L_b) = Ax + L$$

where  $V$  is the residuals;  $A = \partial F / \partial X_a|_{X_a=X_o}$  is the matrix of the partials evaluated with respect to the a priori value  $X_o$  of the parameters;  $L = L_o - L_b$ , where  $L_o = F(X_o)$ ; and  $x = X_a - X_o$ . The least squares estimate of  $x$  is

$$\hat{x} = -(A^T P A)^{-1} A^T P L \quad (A-6)$$

with the corresponding covariance matrix

$$\Sigma_x = (A^T P A)^{-1} \quad (A-7)$$

If a priori knowledge of the covariance matrix corresponding to the parameters exists, then a constrained least squares solution of the parameters is given by

$$\hat{x} = -(P_{ox} + A^T P A)^{-1} A^T P L \quad (A-8)$$

and

$$\Sigma_x = (P_{ox} + A^T P A)^{-1} \quad (A-9)$$

where the parameters are constrained by their a priori weight  $P_{ox}$ .

The partial matrix  $A$  of Eq. (A-3) with respect to the seven transformation parameters, when evaluated with an a priori estimate of the parameters as zero, would become

$$A_{(3n,7)} = \begin{bmatrix} 1 & 0 & 0 & 0 & -Z_{1Ref2} & Y_{1Ref2} & X_{1Ref2} \\ 0 & 1 & 0 & Z_{1Ref2} & 0 & -X_{1Ref2} & Y_{1Ref2} \\ 0 & 0 & 1 & -Y_{1Ref2} & X_{1Ref2} & 0 & Z_{1Ref2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & -Z_{nRef2} & Y_{nRef2} & X_{nRef2} \\ 0 & 1 & 0 & Z_{nRef2} & 0 & -X_{nRef2} & Y_{nRef2} \\ 0 & 0 & 1 & -Y_{nRef2} & X_{nRef2} & 0 & Z_{nRef2} \end{bmatrix} \quad (A-10)$$

Matrix  $A$  will have dimension  $(3n \times 7)$  for  $n$  participating stations. The weight matrix  $P$  of the observation is a full  $(3n \times 3n)$  matrix. The vector  $\tilde{L}$  is of the form

$$L_{(3n,1)} = [(X_{1Ref2} - X_{1Ref1}), (Y_{1Ref2} - Y_{1Ref1}), (Z_{1Ref2} - Z_{1Ref1}), \dots, \dots, \dots, (X_{nRef2} - X_{nRef1}), (Y_{nRef2} - Y_{nRef1}), (Z_{nRef2} - Z_{nRef1})]^T \quad (A-11)$$

The least squares solution is obtained for the parameter vector

$$x_{(7,1)} = [\Delta X, \Delta Y, \Delta Z, R_x, R_y, R_z, \Delta L]^T \quad (A-12)$$

N91-11975

# Computer Simulations of Ions in Radio-Frequency Traps

A. Williams, J. D. Prestage, L. Maleki,  
J. Djomehri, and E. Harabetian  
Communications Systems Research Section

*The motion of ions in a trapped-ion frequency standard affects the stability of the standard. In order to study the motion and structures of large ion clouds in a radio-frequency (rf) trap, a computer simulation of the system that incorporates the effect of thermal excitation of the ions has been developed. Results are presented from the simulation for cloud sizes up to 512 ions, emphasizing cloud structures in the low-temperature regime.*

## I. Introduction

The development of the trapped-ion frequency standard at the Jet Propulsion Laboratory (JPL) is motivated by the potential of this device to achieve stability exceeding one part in  $10^{17}$ . The basis of the potential for this remarkable stability performance is the isolation of the trapped ions from perturbing influences that diminish stability over averaging intervals greater than several hundred seconds. In conventional frequency standards such as hydrogen masers, these effects include the interaction of the atoms with confining walls and phase-destructive collisions of atoms of the same species that are in different energy states.

The confinement of ions in rf traps (the type employed in the JPL trapped-ion standard) is accomplished by exposing the ions to an oscillating field, resulting in a force that acts to direct the ions toward the center of the trap and away from the trap electrodes. This electromagnetic containment is dynamic in nature, as will be detailed below. Therefore, since a large source of frequency offset is

the second-order Doppler or relativistic time-dilation effect due to the motion of the ions, it is appropriate to study the dynamic effects in the rf trap system.

Ions, of course, do not lend themselves to easy visual observation, so the dynamics of trapped particles are often indirectly studied through the observation of the motion of micron-sized particles in a similar trap [1] or by computer simulations of the confining fields and Coulomb interactions experienced by ions in a trap. The introduction of computer simulations to the field of trapped-ion dynamics brings with it a new set of results, which are not from a laboratory experiment but from "computational" studies of a model. Naturally, if the model is an accurate representation of the real system at hand, a computer simulation using that model should duplicate behavior seen in experiments and that predicted by an accurate theory.

This article presents a study that uses computer simulation to examine the motion of ions confined in rf traps to determine possible influences on the frequency stability. While similar studies have been carried out for the case of

a few ions confined in rf traps, the present work includes results for clusters with as many as 512 ions. Thus, this study is a more realistic model for actual frequency standards, where the typical number of confined ions ranges from  $10^3$  to  $10^6$ .

After a brief review of trapped-ion theory in Section II and a description of the simulation in Section III, the computational results are compared with theory in Section IV and with experiment in Section V. In Section VI, the JPL simulation is compared to the results from trapped-ion simulations carried out by other groups, and Section VII outlines the plans for using this simulation to further investigate trapped-ion dynamics.

## II. Trapping Theory

The detailed theory of rf ion traps has been treated previously [2,3,4]. An outline of the theory is presented here to support the discussion of the computer model.

A Paul-type rf trap consists of three hyperbolic electrodes, as shown in Figs. 1(a) and (b).<sup>1</sup> Typically, these electrodes are arranged so that the radius of the trap cavity  $r_0$  is  $\sqrt{2}$  times the height of the cavity. An ac voltage with angular frequency  $\Omega$  is applied in series with a dc voltage between the ring electrode and the two endcap electrodes.

Another type of trap, known as the Penning or dc trap, is described later in this article. Since a full development is not necessary in this discussion, the reader is referred to [2] and [3]. The Penning trap uses the same electrode configuration as the Paul trap, but the rf portion of the electric field is replaced by a static magnetic field in the  $z$ -direction. Hence, neither of the fields in this trap is time-varying.

Applying the Paul trap potentials produces a potential inside the trap of the form

$$\phi(r, z) = \frac{V_{DC} + V_{AC} \cos(\Omega t)}{2r_0^2} (r^2 - 2z^2) \quad (1)$$

The electric field components are obtained by differentiating Eq. (1) as

$$\mathbf{E} = - \left( \frac{V_{DC}}{r_0^2} + \frac{V_{AC} \cos(\Omega t)}{r_0^2} \right) (r - 2z) \quad (2)$$

The trapping effect of this field configuration can be understood as follows. A node exists at the center of the trap ( $r = z = 0$ ) where the electric field is zero; thus, there is no force on a charged particle located there. Depending on the sign of the rf term, at any given time the field is pointing toward the node in one direction and away from the node in the other. It can also be seen in Eq. (2) that the field magnitude increases with increasing distance from the node. This behavior of the field in the interior of the trap leads to a net force that, when averaged over one rf cycle, is directed toward the node [1]. This “ponderomotive” [5] force is what draws charged particles toward the node, trapping them inside the electrodes.

The equations of motion of a single ion subject to the force due to the electric field in Eq. (2) can be expressed using the two dimensionless constants

$$a = \frac{4eV_{DC}}{m\Omega^2 r_0^2} \quad q = -\frac{2eV_{AC}}{m\Omega^2 r_0^2} \quad (3)$$

to represent the trapping parameters, and a dimensionless time parameter

$$\tau = \frac{\Omega}{2} t \quad (4)$$

In this formulation, the equations of motion become forms of the Mathieu differential equation

$$\frac{d^2 r}{d\tau^2} = -(a - 2q \cos(2\tau))r \quad (5)$$

for either radial dimension ( $x$  or  $y$ ), and

$$\frac{d^2 z}{d\tau^2} = 2(a - 2q \cos(2\tau))z \quad (6)$$

for the axial dimension  $z$ . In the expressions for  $a$  and  $q$ ,  $e$  and  $m$  denote the charge and mass of the contained particle, respectively.

The solutions of Eqs. (5) and (6) reveal that an ion influenced by the fields of a Paul trap undergoes a fast rf motion with the same frequency  $\Omega$  as the driving ac potential, and a slow “secular” oscillation, which in the one-dimensional case carries it back and forth through the trap node.

When more than one ion of the same charge-to-mass ratio is present in the trap, each ion has its own equation of motion, consisting of a trapping term as above and a

<sup>1</sup> The linear rf trap described in [8] employs fields similar to the Paul trap, but consists of four cylindrical rods rather than the hyperbolic electrodes.

Coulomb term containing the electrostatic interaction of all other ions. The motion in Cartesian coordinates  $x$ ,  $y$ , and  $z$  of singly ionized atoms of mass  $m$  is described by the set of differential equations

$$\frac{d^2 x_i}{d\tau^2} = -(a - 2q \cos(2\tau))x_i - \frac{e^2}{\pi\epsilon_0 m \Omega^2} \sum_j \frac{(x_j - x_i)}{r_{ij}^3} \quad (7)$$

$$\frac{d^2 y_i}{d\tau^2} = -(a - 2q \cos(2\tau))y_i - \frac{e^2}{\pi\epsilon_0 m \Omega^2} \sum_j \frac{(y_j - y_i)}{r_{ij}^3} \quad (8)$$

and

$$\frac{d^2 z_i}{d\tau^2} = 2(a - 2q \cos(2\tau))z_i - \frac{e^2}{\pi\epsilon_0 m \Omega^2} \sum_j \frac{(z_j - z_i)}{r_{ij}^3} \quad (9)$$

Here,  $x_i$ ,  $y_i$ , and  $z_i$  are the Cartesian coordinates of the  $i$ th ion, and  $r_{ij}$  signifies the distance between ions  $i$  and  $j$ .

### III. Model for the Simulation of Trapped-Ion Dynamics

Equations (7), (8), and (9) completely describe the trapping and interparticle Coulomb effects on each ion in a cloud of arbitrary size. Because the interest here is the low-temperature behavior of collections of ions, a damping that is proportional to particle velocity is introduced in order to cool the ions. This adds a final term to the differential equations

$$\begin{aligned} \frac{d^2 x_i}{d\tau^2} = & -(a - 2q \cos(2\tau))x_i \\ & - \frac{e^2}{\pi\epsilon_0 m \Omega^2} \sum_j \frac{(x_j - x_i)}{r_{ij}^3} - \frac{2\mathcal{D}}{\Omega} \frac{dx_i}{d\tau} \end{aligned} \quad (10)$$

$$\begin{aligned} \frac{d^2 y_i}{d\tau^2} = & -(a - 2q \cos(2\tau))y_i \\ & - \frac{e^2}{\pi\epsilon_0 m \Omega^2} \sum_j \frac{(y_j - y_i)}{r_{ij}^3} - \frac{2\mathcal{D}}{\Omega} \frac{dy_i}{d\tau} \end{aligned} \quad (11)$$

and

$$\begin{aligned} \frac{d^2 z_i}{d\tau^2} = & 2(a - 2q \cos(2\tau))z_i \\ & - \frac{e^2}{\pi\epsilon_0 m \Omega^2} \sum_j \frac{(z_j - z_i)}{r_{ij}^3} - \frac{2\mathcal{D}}{\Omega} \frac{dz_i}{d\tau} \end{aligned} \quad (12)$$

where  $\mathcal{D}$  is a coefficient of the damping. Once again, the factor of  $2/\Omega$  in the third term of each of these equations appears because of the change of variable to dimensionless time  $\tau$ . The inverse of the damping coefficient  $1/\mathcal{D}$  is known as the damping time  $\tau_d$  of a system with damping  $\mathcal{D}$ .

In a manner similar to the case of particles undergoing Brownian motion, a random walk in velocity is introduced, which simulates the process that sustains ion motion in the presence of damping, thus preventing the temperature from reaching zero. The growth in kinetic energy stemming from this random-walk step in velocity is balanced by the velocity damping introduced in the above equations of motion. This balance determines the size of the random-walk step that yields the desired final temperature.

The signature of a random-walk progression is that the mean square of the quantity in question increases linearly with the number of steps  $\mathcal{N}$  in the walk [6]. Here, this quantity will be a velocity in the space of the dimensionless parameter  $\tau$ , denoted as

$$v_\tau = \frac{dx}{d\tau}$$

Thus, the magnitude of the random-walk step in the dimensionless time space  $\Delta v_\tau$  must obey

$$\langle v_\tau^2 \rangle = \mathcal{N}(\Delta v_\tau)^2 \quad (13)$$

or in terms of the usual velocity  $v$ ,

$$\left\langle \left( \frac{dx}{d\tau} \right)^2 \right\rangle = \frac{4}{\Omega^2} \left\langle \left( \frac{dx}{dt} \right)^2 \right\rangle \quad (14)$$

$$= \frac{4}{\Omega^2} \langle v^2 \rangle = \mathcal{N}(\Delta v_\tau)^2 \quad (15)$$

Introducing a short dimensionless time interval  $h$  between random-walk steps so the  $\mathcal{N}$  steps of length  $h$  represent a span of dimensionless time  $\tau$ ,

$$\frac{4}{\Omega^2} \langle v^2 \rangle = \frac{\mathcal{N}h}{h} (\Delta v_\tau)^2 = \tau \frac{(\Delta v)^2}{h} \quad (16)$$

and, hence,

$$\left( \frac{d\langle v^2 \rangle}{d\tau} \right)_{rw} = \frac{\Omega^2 (\Delta v)^2}{4h} \quad (17)$$

where  $rw$  suggests that this is due to the random walk. This is no more than a restatement of the fact that the mean square of the velocity grows linearly in time in the random walk. The growth of  $\langle v^2 \rangle$  due to the random walk is balanced by the damping of velocity per the damping terms in Eqs. (10), (11), and (12). The term due to damping becomes

$$\frac{d\langle v^2 \rangle}{d\tau} = -\frac{4}{\Omega} \mathcal{D} v^2 \quad (18)$$

For this damping term, it can be shown that when the interval over which  $v^2$  is averaged is small in comparison to  $\tau_d/2$ ,  $\langle v^2 \rangle = v^2$ . Then

$$\begin{aligned} \left( \frac{d\langle v^2 \rangle}{d\tau} \right)_{damp} &= \left( \frac{d(v^2)}{d\tau} \right)_{damp} \\ &= -\frac{4}{\Omega} \mathcal{D} \langle v^2 \rangle \end{aligned} \quad (19)$$

Setting the sum of the two terms to zero to balance the effects yields

$$\begin{aligned} \frac{d\langle v^2 \rangle}{d\tau} &= \left( \frac{d\langle v^2 \rangle}{d\tau} \right)_{rw} + \left( \frac{d\langle v^2 \rangle}{d\tau} \right)_{damp} \\ &= \frac{\Omega^2}{4} \frac{(\Delta v_\tau)^2}{h} - \frac{4}{\Omega} \mathcal{D} \langle v^2 \rangle = 0 \end{aligned} \quad (20)$$

To get a random direction of the change in velocity, a separate random walk is taken in each of the Cartesian directions  $x$ ,  $y$ , and  $z$ . Therefore, the  $\Delta v_\tau$  in Eq. (20) is the size of the step in any one of these dimensions, and  $v^2$  is the velocity squared in any one dimension. The temperature  $T$  of the system is introduced by use of the equipartition of energy

$$\frac{1}{2} k_B T = \frac{1}{2} m \langle v^2 \rangle$$

Making this substitution for  $\langle v^2 \rangle$  and solving for  $\Delta v_\tau$  in Eq. (20) yields the magnitude of the random-walk step in one dimension:

$$\Delta v_\tau = \sqrt{\frac{16 \mathcal{D} h k_B T}{m \Omega^3}} \quad (21)$$

The sign of this velocity boost is determined by using a random-number generator that chooses each sign 50 percent of the time.

The approach to simulating a system described by these equations is obviously to determine the time evolution of the position-momentum states of the particles. The necessity of a constant small time step  $h$  in the random-walk formulation of temperature creates an ideal situation in which to use a fourth-order Runge-Kutta integration, with a constant time step, of the differential Eqs. (10), (11), and (12). The time between the random-walk steps  $h$  will also be the time step of the integration routine, so that the random walk can be executed at each integration step. The only requirement on the size of  $h$  is that it is small enough so that the errors (of the order  $h^5$ ) are “small enough.” In the case that the Runge-Kutta integration is calculating a specific numerical solution, one checks the solution by cutting  $h$  in half and checking that the two solutions agree [7]. In the case described here, where the integration is producing the motion of the ions in time, the appropriateness of the size of  $h$  can be checked by determining that halving  $h$  does not produce any noticeable increase in smoothness of the motion. In the cases examined, it was found that an integration step of  $1/20$  the period of the rf-driven motion,  $h = 2\pi/20\Omega$ , gives the desired precision.

The integration will give the three components of position and velocity for each particle at each step of the integration. This position-momentum state contains all the information about the system; other descriptions of the system are just obtained from the position-momentum state as appropriate. For instance, to obtain a kinetic energy for the entire cloud at any given time, the velocities of each particle are used to compute  $mv^2/2$ , and summing this quantity over each particle in the cloud gives an instantaneous total kinetic energy.

## IV. Comparison of Results of the Model With Theory

Verification that the modeling of trapping voltages and Coulomb interactions is an accurate representation of a system of ions in a trap can be performed by checking the behavior of ions in situations readily calculable from the equations of motion. For instance, it was previously mentioned that an ion in the rf field oscillates with a slow secular motion in addition to the fast driven motion (the “micromotion”). The secular frequencies are given by [8] as

$$\omega_r^2 = \frac{e^2 V_{AC}^2}{2m^2 \Omega^2 r_0^4} + \frac{e V_{DC}}{m r_0^2} \quad (22)$$

and

$$\omega_z^2 = \frac{2e^2 V_{AC}^2}{m^2 \Omega^2 r_0^4} - \frac{2e V_{DC}}{m r_0^2} \quad (23)$$

These quantities are easily calculated for any combination of the trapping voltages. In particular, if either  $V_{AC}$  or  $V_{DC}$  is zero, the expressions for secular frequencies reduce to a single term. Another simple case for which this happens is the so-called “spherical” case often employed in trap operation, in which  $\omega_r^2$  and  $\omega_z^2$  are chosen to be equal by setting

$$V_{DC(sph)} = \frac{e V_{AC}^2}{2m \Omega^2 r_0^2} \quad (24)$$

which gives

$$\omega_r^2 = \omega_z^2 = \frac{e^2 V_{AC}^2}{m^2 \Omega^2 r_0^4} \quad (25)$$

To isolate the trapping and Coulomb forces, the parameters  $\mathcal{D}$  and  $T$  in Eqs. (10), (11), (12), and (21) are simply set to zero. Then a simulation of a single ion in the trap clearly shows the micromotion (if  $V_{AC} \neq 0$ ) and the slower secular oscillation, as in Fig. 2. The period of the oscillation can then be found by measuring the time between three successive zero-crossings of the position versus time graph. Since the position data exist only at each integration step  $h$ , the location in time of a zero-crossing is determined only to the accuracy of  $\pm \frac{1}{2}h$ .

The values obtained from the simulation of the periods  $T$  corresponding to these frequencies and the corresponding values from the above theory are summarized in Table 1. The agreement of the periods and the smooth sinusoidal motion seen in Fig. 2 indicate that the integration of the trapping and Coulomb terms is working properly.

If the temperature parameter accurately models ordinary temperature in kelvins, then, as previously suggested, the particles should have a thermal energy of  $\frac{1}{2}k_B T$  per degree of freedom. This can be checked in the simulation by computing the mean-square velocity of the guiding center of motion, that is, a  $\langle v^2 \rangle$  that does not include the driven micromotion of the particles. This is done by computing the motion in time of the average position of each particle over a micromotion period, and then averaging the square of this quantity over all ions and over a satisfactory span of time.

For various degrees of damping and number of ions  $N_{ions}$ , the computational value of  $v^2$  is in agreement with the theoretical value given by the equipartition of energy,

as summarized in Table 2 for  $T = 0.005$  K and  $m = 137$  atomic mass unit (amu). The uncertainties reported are computed as the standard deviation in the  $N$  measurements.

The theoretical value is within the uncertainty range of each of the computational values. This indicates that a random walk in velocity space is a valid model of temperature.

## V. Comparison of the Results of the Model With Experiments

Another way to examine the validity of using this simulation is to compare the ion dynamics with results observed by experimenters who have trapped charged micron-sized particles in an rf trap. Using a suitable magnification method, these particles can be observed visually or photographed. Thirty-two particles trapped with an rf field only ( $V_{DC} = 0$ ) have been seen to have a “crystalline” state where the particles arrange themselves in three concentric circles containing 4, 14, and 14 particles [4]. This photograph is shown in Fig. 3(a); the appearance of each particle as a line is due to the fast micromotion. Simulating this situation gives a similar “photograph” when the  $x$ - $y$  positions of all particles are plotted over one period of the micromotion, Fig. 3(b).

When  $V_{DC} \neq 0$  and there are many ( $\approx 100$ ) charged particles in a dust trap, experimenters have seen a characteristic picture in which the large micromotion of the particles farthest away from the node maps out the electric field pattern of the trap, as shown in Fig. 4(a) [4]. In a simulation of 128 particles, the plot of their motions in three dimensions yields a similar picture, Fig. 4(b).

Experiments in which small numbers of ions in an rf trap are caused to visibly fluoresce have been carried out [9]. These experiments indicate that with  $V_{DC} = 0$ , a pair of ions has a stable configuration in which the pair is situated symmetrically about the node in the  $z = 0$  plane, Fig. 5(a). With  $V_{DC} > V_{DC(sph)}$ , however, the  $z$  direction is preferred, and the most stable configuration is with the pair centered about the node in the  $r = 0$  plane, Fig. 5(b). These results may be understood from the use of a harmonic pseudopotential that is derived through the force on a trapped particle averaged over one rf cycle [2]: in the first case, with  $V_{DC} = 0$ , the restoring force in the pseudopotential well is four times as strong in the  $r$  direction as in the  $z$  direction. However, in the case that  $V_{DC} > V_{DC(sph)}$ , the force is stronger in the axial ( $z$ ) direction.

Duplicating these situations in the simulation produces the same most stable configuration in each case. Furthermore, the nature of the potential well is evident in noting the progress of the system toward this configuration when the initial conditions are far from it. For instance, the ions may be started on the  $z = 0$  plane, with the voltages set in the  $V_{DC} > V_{DC(sph)}$  case. Note that by Eq. (2), there is no trapping force in the  $z$  direction on a particle situated at  $z = 0$ . Hence, the trapping force alone would never bring the particles out of the  $z = 0$  plane no matter how deep the well in the  $z$  direction! In the actual system, of course, collisions with the atoms of the buffer gas will knock the ions off the plane, allowing the force to have nonzero components in both directions. The ions can then drop into the well.

It is evident that in a computer model, some noise that can mix the  $r$  and  $z$  directions in this way must be included [10]; in the simulation described here, the random walk in velocity provides this noise. Starting the ions centered about the node on the  $z = 0$  plane produces motion as shown in Fig. 6, which is obtained by plotting the positions of the two ions when they are near the maxima of their secular oscillation for several secular periods. Initially the ions oscillate only in the  $r$  direction; then the random-walk step moves the ions off the  $z = 0$  plane. Note the symmetry of the system: the ions stay opposite one another as if connected by a spring, as suggested by the dashed lines between the pairs of ions. Now that the trapping force has both  $r$  and  $z$  components, the ions can move toward their most stable configuration along the  $z$  axis. They “overshoot” the  $z$  axis due to the angular momentum they have from being drawn toward the axis, and then return, precessing about the axis in this manner (suggested by the solid arrows) until the motion is damped out and the ions sit quite stably along the axis with only slight vibrations around their equilibrium positions (Fig. 7).

## VI. Comparison to Other Simulations

Although there are no other large-scale simulations of rf traps, simulations [12] of large numbers of ions in a Penning trap have been conducted. These are based on the notion that, disregarding the motion due to the confining fields, confined ions are regarded as thermal particles with a mutual Coulomb repulsion in either type of trap. This repulsion keeps the ions quite far apart from one another, similar to states observed in a one-component plasma [11].

A useful way to describe the energy state of a plasma is by its coupling constant  $\Gamma$ , which is the ratio of Coulomb energy to thermal kinetic energy, or  $e^2/ak_B T$ . Here  $a$  is

an average distance between the ions,  $k_B$  is the Boltzmann constant, and  $T$  is temperature.

If  $\Gamma$  is  $\ll 1$ , the thermal energy is much greater than the Coulomb energy, and the ions may be expected to move quite freely as thermal particles without much regard to the comparatively weaker Coulomb force. However, if  $\Gamma \gg 1$ , the Coulomb energy exceeds the kinetic energy, and less thermal motion would be expected in the ion system. The theory predicts a solid-like state with very little thermal motion for  $\Gamma \geq 180$ , and a liquid-like state with an intermediate level of thermal diffusion for  $2 \leq \Gamma \leq 180$  [5].

The Penning trap simulations have shown that in the spherical case, when the coupling  $\Gamma$  between the ions is large, the ions form into concentric spherical shells centered on the trap node. Qualitative agreement, with the exception in some cases of the presence of cylindrical shells rather than spheroids, is seen in actual experiments on many Penning-trapped ions, which can be visually observed by causing the ions to fluoresce [13].

The simulation described here has also shown this behavior for ion clouds of up to 512 ions. This is shown in Figs. 8(a), (b), and (c) for several cloud sizes by plotting the positions of the ions in two dimensions as  $|z|$  versus  $\sqrt{x^2 + y^2}$ . This picture has the effect of sweeping around all  $\phi$  (the longitudinal angle in spherical coordinates) and folding into one quadrant. Using this representation, the adding of additional shells with increasing numbers of ions due to their Coulomb repulsion is clear: 128 ions arrange themselves in three shells, 256 ions in four shells, and for 512 ions a fifth shell is necessary. By taking the “photograph” during the crystallization process, information is gained on the process itself. It is seen in Fig. 9 that in this 512-ion case, the ion density begins to build up on what will be the fifth (outer) and the fourth shells in Fig. 8(b) while the ions are still mixing with respect to one another.

It has been reported that for very large  $\Gamma$ , the ions on the outer shell of a large ion cloud will have very little thermal motion, in accordance with the strongly coupled plasma theory. Furthermore, these ions arrange themselves into a two-dimensional hexagonal lattice on the surface of the shell [12]. The simulation described here also produces this result in both 128- and 256-ion clouds. The lattice itself can be examined in a polar plot ( $\theta$  versus  $\phi$ ) as in Fig. 10(a), or in a “time-lapse” three-dimensional plot of ions on the hemisphere, Fig. 10(b).

The lattice can also be examined in a more quantitative fashion by considering the spatial correlation of the ions on the outer shell. The coordination number function  $C(s)$

describes the average number of ions within a distance  $s$  of one ion, and the correlation function  $c(s)$  is given in terms of  $C(s)$  as [12]

$$\int_0^s 2\pi s c(s) ds = C(s) \quad (26)$$

A regular two-dimensional hexagonal lattice has six nearest neighbors to any one point, and eighteen next-nearest neighbors, so the function  $C(s)$  when plotted against  $s$  should show horizontal shoulders at  $C(s) = 6$  and  $C(s) = 18$ . The correlation function can be obtained from  $C(s)$  by Eq. (26).

For the 128-ion cloud in Figs. 10(a)–10(d), which is at  $\Gamma = 556$ , the highly regular hexagonal lattice is evident in the polar plot and the three-dimensional plot, and supported by the shape of  $C(s)$ , Fig. 10(c). The correlation function itself,  $c(s)$ , has three smooth, well-defined peaks, Fig. 10(d), which also indicates a large degree of order.

At  $\Gamma = 139$ , some changes in the lattice are evident in the three-dimensional plot (Fig. 11). This time-lapse plot shows that over a length of time, the ions wander quite a bit more from their positions in the hexagonal lattice, but there is still a fair degree of order and there is not much exchanging of positions between ions.

When  $\Gamma$  is lowered to 65, the time-lapse plot shows that the ions are wandering quite randomly over the surface of the sphere, Fig. 12(a). Indeed, the second shoulder of  $C(s)$  has been virtually wiped out, Fig. 12(b), indicating that order exists only on the scale of the first unit cell. The plot of  $c(s)$  has similarly degraded, Fig. 12(c). The simulations

that have been done on Penning trap systems also show this effect, in which the ions are crystallized into shells but diffuse freely over the surface of the shell, for low (but  $> 1$ ) values of  $\Gamma$  [12].

## VII. Conclusions

It has been shown that this model of the Paul trap system gives results that agree with trapping theory, visual observations of trapped particles, and other computer simulations of trapped charged particles. Armed with this model, questions that are important to the trapped-ion frequency standard can now be examined, such as the effect of “rf heating,” in which a large ion cloud may gain extra energy from the effects of the trapping rf field [14]. Clearly, the normal modes of oscillation of a system of  $N$  ions in this rf quadrupole field have not been solved analytically; the simulation may be an excellent way to understand the oscillations of such a large system. Also, by simply revising the differential equations in the computer program, rf simulations of other trap geometries (such as the linear ion trap described in [8], which is now in use) can be done.

To the authors’ knowledge, there have been no simulations of ions in such a linear trap, and although experiments have been done at JPL on the confinement of micron-sized particles in a linear trap, they are not nearly as detailed as the studies in [4], and little is known about the crystalline structures that may form in a linear trap. Therefore, it is indeed important to have confidence in the integrity of the simulation by examining a system that has been widely studied (the hyperbolic trap) before progressing to less understood systems.

## Acknowledgments

The authors wish to acknowledge the use of the JPL and NASA Ames Numerical Aerodynamic Simulation supercomputers to perform these simulations.

## References

- [1] A. Williams, R. Melbourne, L. Maleki, G. Janik, and J. Prestage, "An Apparatus for the Electrodynamical Containment of Charged Macroparticles," *TDA Progress Report 42-98*, vol. April-June 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 57-62, August 15, 1989.
- [2] H. G. Dehmelt, *Advanced Atomic Molecular Physics*, Vol. III: New York: Academic Press, 1967.
- [3] D. J. Wineland, "Spectroscopy of Stored Ions," *Precision Measurement and Fundamental Constants II*, U. S. National Bureau of Standards, Special Publication 617, pp. 83-92, Boulder, Colorado, 1984.
- [4] R. F. Wuerker, H. Shelton, and R. V. Langmuir, "Electrodynamical Containment of Charged Particles," *J. Appl. Phys.*, vol. 30, no. 3, pp. 342-349, March 1959.
- [5] J. J. Bollinger and D. J. Wineland, "Microplasmas," *Scientific American*, vol. 262, no. 1, pp. 124-130, January 1990.
- [6] P. M. Morse, *Thermal Physics*, New York: W. A. Benjamin, Inc., 1969.
- [7] G. Arfken, *Mathematical Methods for Physicists*, Orlando, Florida: Academic Press, Inc., 1985.
- [8] J. D. Prestage, G. J. Dick, and L. Maleki, "New Ion Trap for Frequency Standard Applications," *J. Appl. Phys.*, vol. 66, no. 3, pp. 1013-1017, August 1, 1989.
- [9] D. J. Wineland, J. C. Bergquist, Wayne M. Itano, J. J. Bollinger, and C. H. Manney, "Atomic-Ion Coulomb Clusters in an Ion Trap," *Phys. Rev. Lett.*, vol. 59, no. 26, pp. 2935-2938, December 28, 1987.
- [10] J. Hoffnagle, R. G. DeVoe, L. Reyna, and R. G. Brewer, "Order-Chaos Transition of Two Trapped Ions," *Phys. Rev. Lett.*, vol. 61, no. 3, pp. 255-258, July 18, 1988.
- [11] J. H. Malmberg and T. M. O'Neil, "Pure Electron Plasma, Liquid, and Crystal," *Phys. Rev. Lett.*, vol. 39, no. 21, pp. 1333-1336, November 21, 1977.
- [12] D. H. E. Dubin and T. M. O'Neil, "Computer Simulation of Ion Clouds in a Penning Trap," *Phys. Rev. Lett.*, vol. 60, no. 6, pp. 511-514, February 8, 1988.
- [13] S. L. Gilbert, J. J. Bollinger, and D. J. Wineland, "Shell-Structure Phase of Magnetically Confined Strongly Coupled Plasmas," *Phys. Rev. Lett.*, vol. 60, no. 20, pp. 2022-2025, May 16, 1988.
- [14] H. Dehmelt, in *Advances in Laser Spectroscopy*, F. T. Arecchi, F. Strumia, and H. Walther, editors, New York: Plenum, pp. 153-188, 1983.

**Table 1. Comparison of secular oscillation periods**

$V_{AC}$	$V_{DC}$	Computational $T_r$	Theoretical $T_r$	Computational $T_z$	Theoretical $T_z$
120	0	$1048 \pm 1$	1048.0	$523 \pm 1$	524.0
0	40	$211 \pm 1$	210.9	$149 \pm 1$	149.1
120	$V_{DC(sph)}$	$740 \pm 1$	741.0	$740 \pm 1$	741.0

Note: Voltages are reported in volts, and periods in integration steps,  $h$ .

**Table 2. Comparison of thermal energies**

$N_{ions}$	$\tau_d/\tau_s$	Computational $v^2$	Theoretical $v^2$	$N$
128	9.77	$0.30079 \pm 0.01475$	0.301231	24
128	26.6	$0.27942 \pm 0.02556$	0.301231	26
256	4.9	$0.29976 \pm 0.00576$	0.301231	19

Note: The value  $\tau_d/\tau_s$  is a measure of the relative amount of damping: the ratio of the damping time to the single-particle secular period.

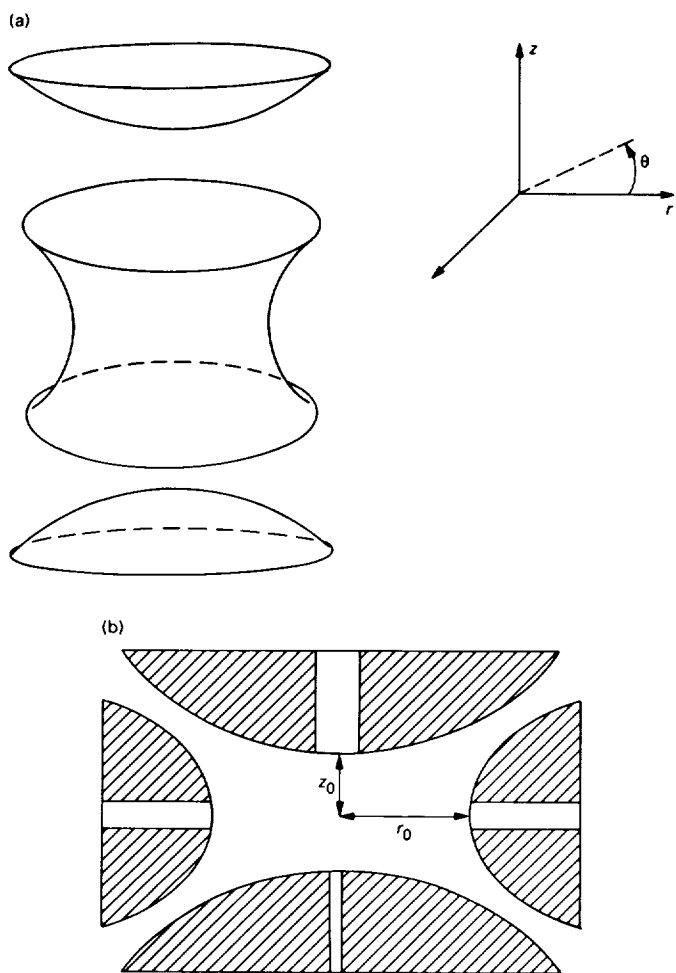


Fig. 1. A hyperbolic trap: (a) the three electrodes, consisting of one ring electrode and two endcaps, and (b) the assembled hyperbolic trap in cross section.

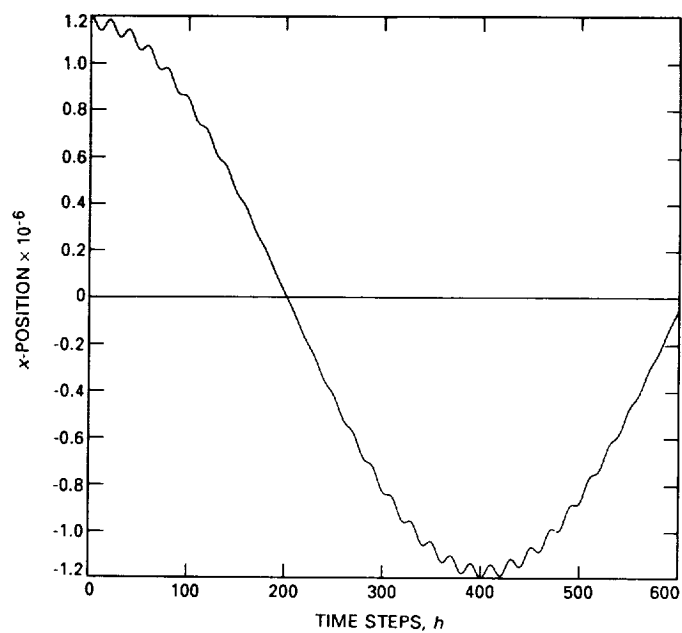


Fig. 2. One-dimensional single-particle trajectory in a Paul trap, showing the fast driven micromotion superimposed on the slower secular motion.

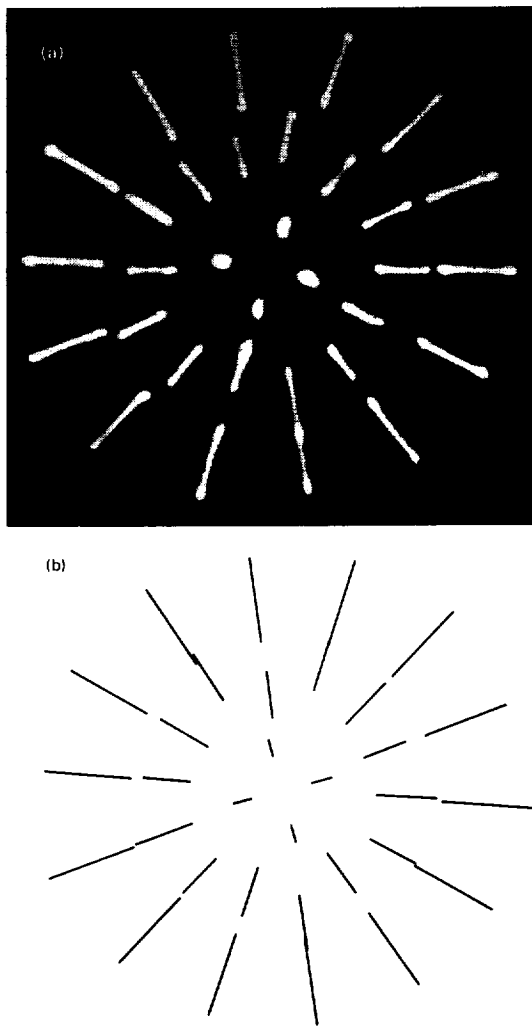


Fig. 3. Comparison of rf-trapped macroparticles with simulation: (a) photograph from [4] showing the "crystalline" state of 32 particles, and (b) results from the simulation of 32 trapped particles.

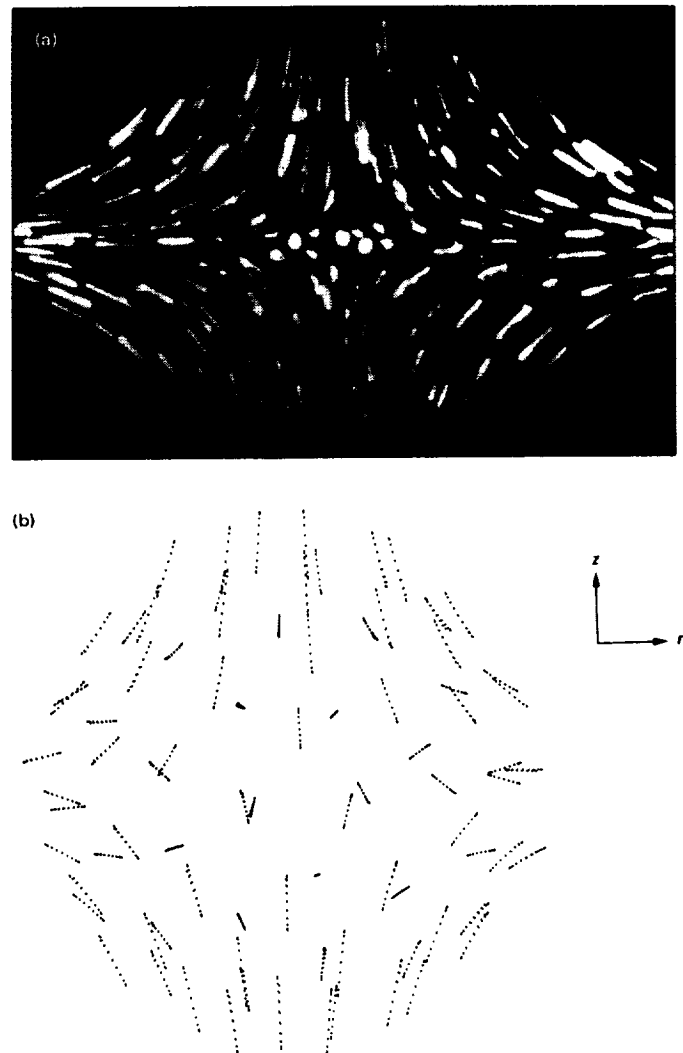


Fig. 4. Comparison of  $\approx 100$  rf-trapped particles with simulation: (a) photograph from [4] showing  $\approx 100$  particles outlining the fields in the trap, and (b) results from the simulation of 128 rf-trapped particles.

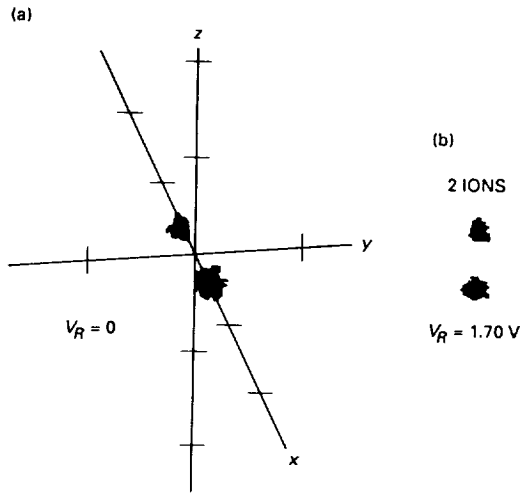


Fig. 5. Photograph from [9] showing two trapped ions whose preferred configuration is to be situated (a) symmetrically about the node in the radial direction, and (b) symmetrically about the node in the axial direction.

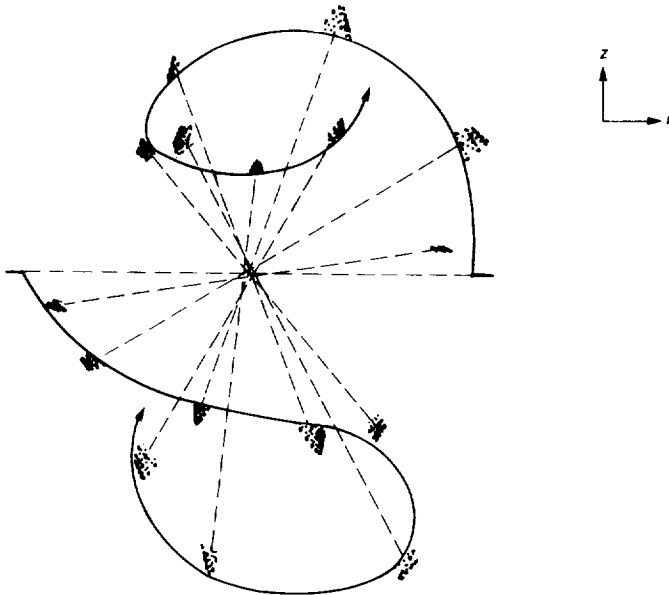


Fig. 6. Simulation of a pair of ions whose preferred configuration is to lie on the z-axis; when placed on a radial axis, they are drawn toward the vertical and precess about the z-axis.

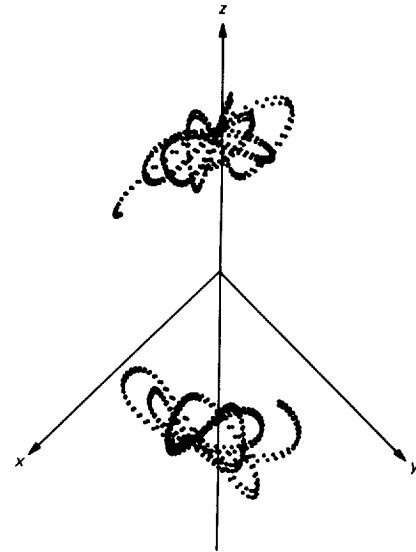


Fig. 7. The system of Fig. 6, after much of the precessional motion has damped out, sits stably along the z-axis.

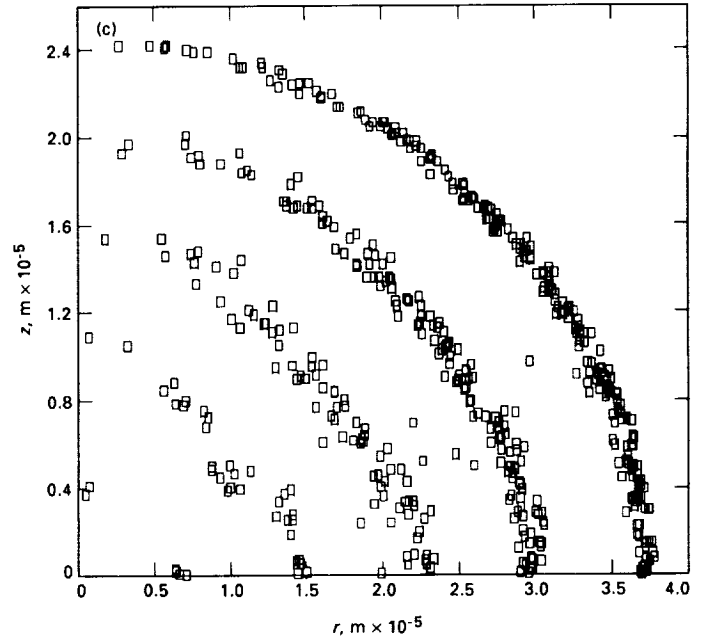
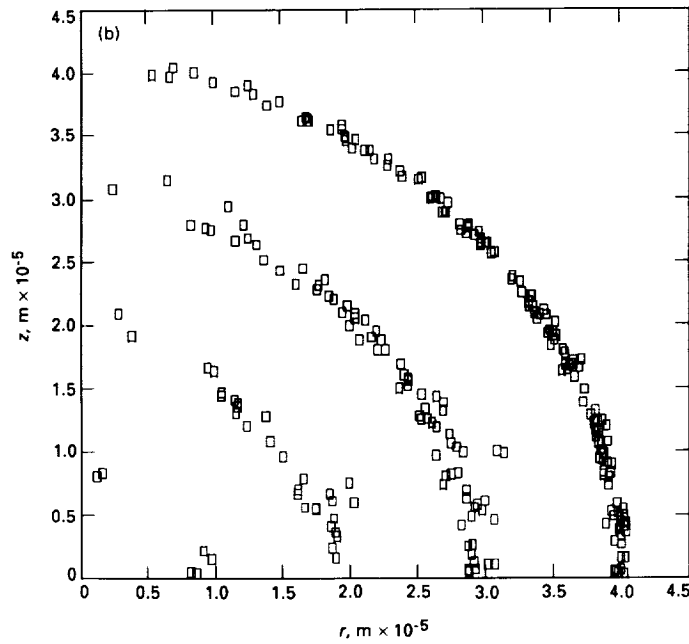
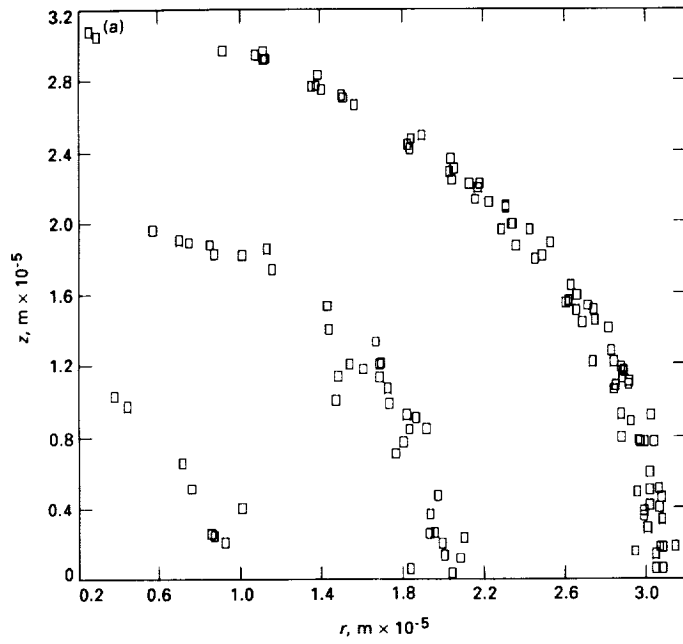
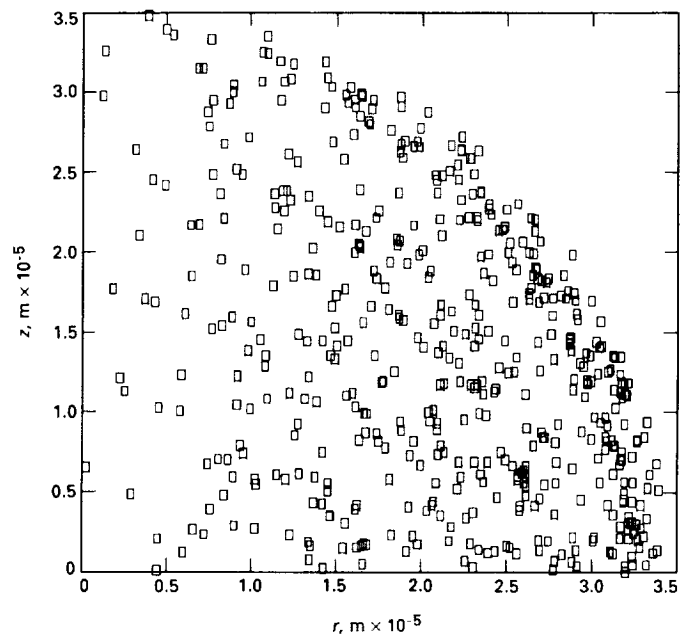


Fig. 8. An  $r$  versus  $z$  plot of the positions of ions in a simulation: (a) 128 ions form three concentric spherical shells; (b) 256 ions form four spherical shells; and (c) 512 ions form five spherical shells.



**Fig. 9.** An  $r$  versus  $z$  plot of a simulation of 512 ions during their crystallization, showing the buildup of density where the fourth and fifth shells are forming.

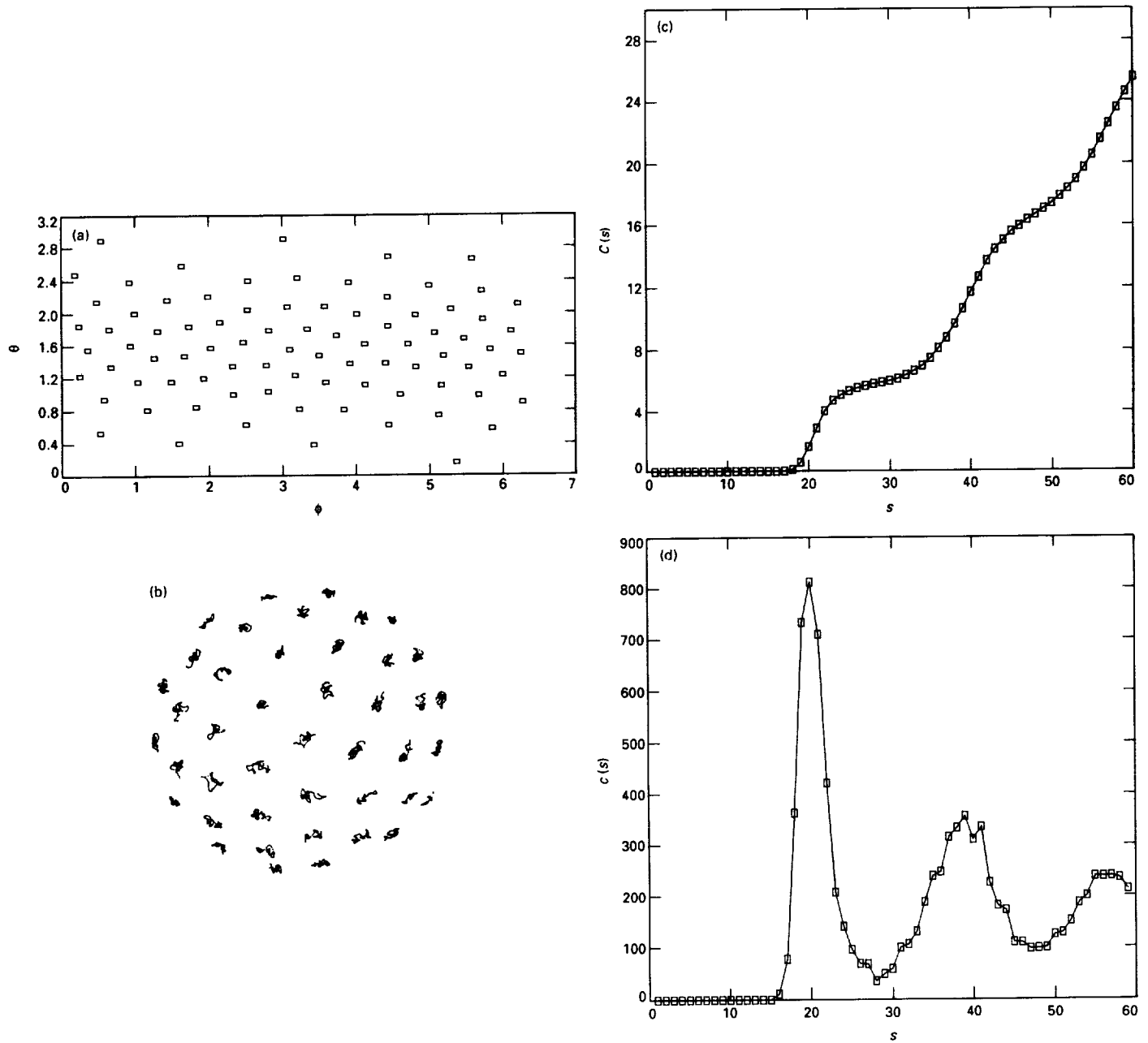


Fig. 10. A 128-ion cloud: (a) polar ( $\theta$  versus  $\phi$ ) plot of ions on the outer shell at  $\Gamma = 556$ , showing the two-dimensional hexagonal lattice that forms on the surface of the shell; (b) three-dimensional view of ions on the near hemisphere of the outer shell (the positions on this "time-lapse" plot are sampled once per micromotion cycle); (c) the coordination number function  $C(s)$  of the ions in the system of (a), showing shoulders at  $C(s) = 6$  and  $C(s) = 18$ , indicating a hexagonal lattice; and (d) the correlation function  $c(s)$  of this system. The three smooth, well-defined oscillations of this function indicate a high degree of order.

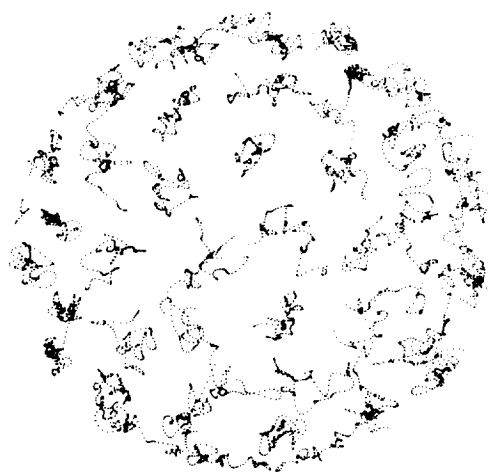


Fig. 11. Time-lapse plot similar to Fig. 10(b), showing the outer hemisphere of a 128-lon cloud at  $\Gamma = 139$ . More thermal motion is seen here than in Fig. 10(b).

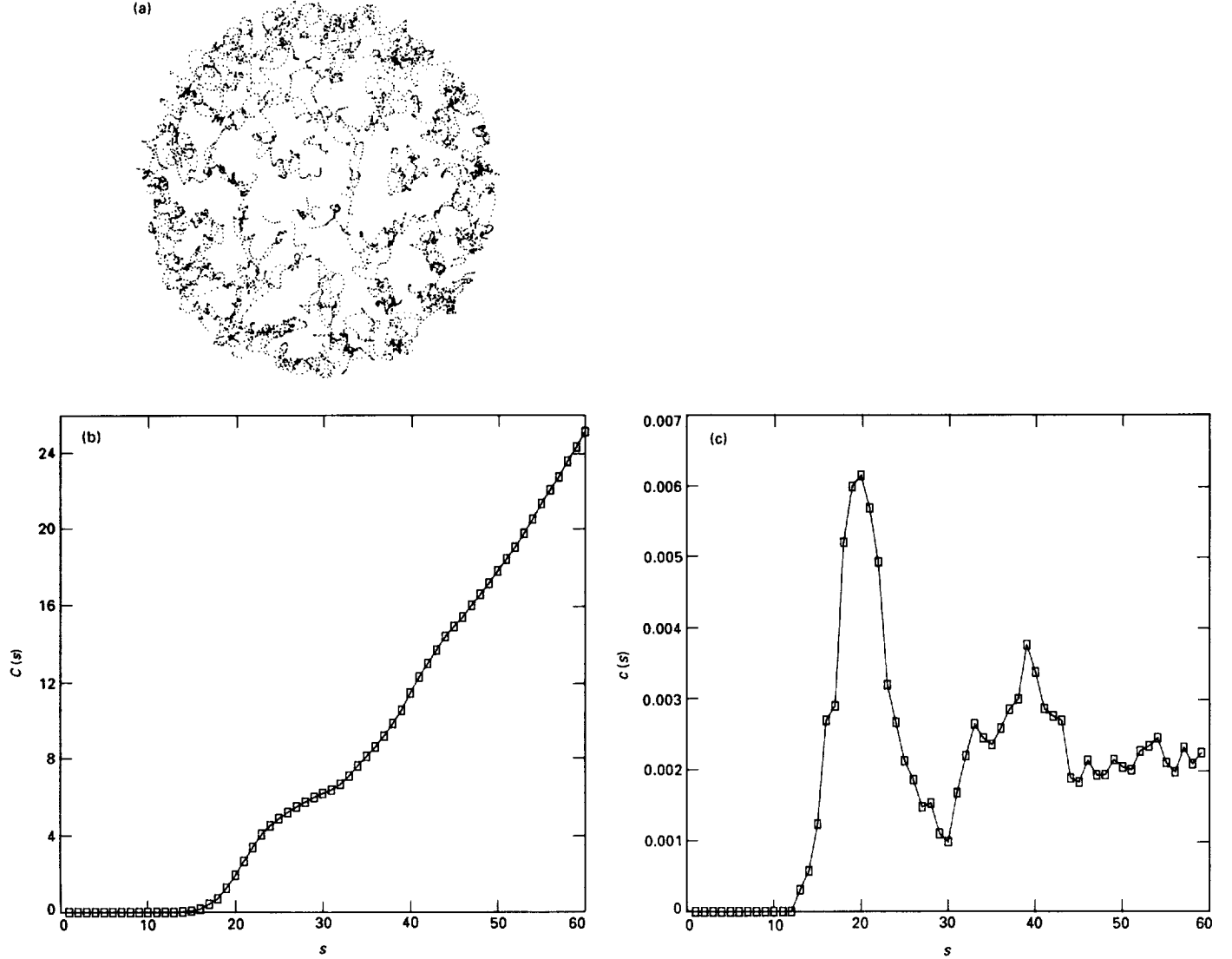


Fig. 12. A 128-ion cloud at  $\Gamma = 65$ : (a) time-lapse plot showing that the ions diffuse about the spherical shell much more in this case than at the higher values of  $\Gamma$ ; (b) coordination number function  $C(s)$  of the system of (a), showing that in comparison to Fig. 10(c) the shoulder at  $C(s) \approx 18$  has disappeared, indicating that at this lower  $\Gamma$  order exists over length scales of a one-unit hexagonal cell only; and (c) correlation function  $c(s)$  of this system. In comparison to the  $c(s)$  in Fig. 10(d), this one is far noisier and lacks the third oscillation, indicating less order in this low- $\Gamma$  system.

N91-11976

# Design of a Fiber-Optic Transmitter for Microwave Analog Transmission With High Phase Stability

R. T. Logan, Jr., G. F. Lutes, L. E. Primas, and L. Maleki  
Communications Systems Research Section

*The principal considerations in the design of fiber-optic transmitters for highly phase-stable radio frequency and microwave analog transmission are discussed. Criteria for a fiber-optic transmitter design with improved amplitude and phase-noise performance are developed through consideration of factors affecting the phase noise, including low-frequency laser-bias supply noise, the magnitude and proximity of external reflections into the laser, and temperature excursions of the laser-transmitter package.*

## I. Introduction and Overview

Phase-stable transmission of analog signals is crucial for ultrastable frequency-reference distribution, phased-array antenna synchronization systems, and radio science experiments. Better phase-noise performance from directly modulated semiconductor laser-diode fiber-optic systems is compulsory for precise distribution of next-generation frequency references with frequency stability of  $10^{-15}/\tau$  [1], and for improved synchronization of large phased-array antenna systems.

It has been shown [2] that the phase noise of a radio-frequency (RF) signal transmitted over a fiber-optic link via direct modulation of a semiconductor laser diode is equal to the signal-to-noise ratio (SNR) of the link, determined by the intensity noise of the laser diode. While this is true at high offset frequencies ( $f \geq 1$  MHz) from the RF carrier, the results of this study indicate that RF sig-

nals transmitted through a fiber-optic link by direct (current) modulation of a semiconductor laser diode acquire close-to-carrier phase noise (from 0.1 Hz to 10 kHz offsets) which can be much larger than the SNR of the link. Thus, it is noted that the recent dramatic improvements in the high-frequency noise characteristics of semiconductor laser diodes have not produced commensurate improvements in close-to-carrier RF phase-noise performance. Also, it appears that close-to-carrier RF phase noise is not determined by the relative intensity noise (RIN) of the laser in the modulation band, since lasers with widely varying levels of RIN have been found to yield similar close-to-carrier phase-noise performance.

In this article, characterizations of the close-to-carrier RF phase noise of fiber-optic links employing both Fabry-Perot (FP) and distributed-feedback- (DFB-) type laser diode transmitters are presented, as well as a discussion

of the effects of bias current fluctuations, optical reflections, and temperature variations on the magnitude of the RF phase noise. From these findings, a set of specific criteria have emerged for the design of fiber-optic links with improved RF phase-noise performance.

## II. Characterizations of Additive RF Phase Noise

The additive RF phase noise of two different fiber-optic links was measured using the phase-detector technique [3] driven by a 100-MHz signal from a hydrogen-maser frequency standard. The measurement system is illustrated in Fig. 1. The results of phase-noise measurements of fiber-optic systems employing both a multi-longitudinal-mode FP laser diode [4] and a single-longitudinal-mode DFB laser diode [5] are presented in Fig. 2. Both systems were operated with a 4-km length of fiber and low-reflection expanded-beam optical connectors. The DFB package incorporated a 30-dB internal optical isolator. The most surprising result is that the phase noise of the isolated DFB transmitter at 1 Hz is identical to that of the non-isolated FP transmitter, even though the SNR of the DFB laser is 15 to 20 dB better than the FP laser at 100 MHz.

Initially, it was suspected that the close-to-carrier phase-noise floor would be set by the magnitude of the low-frequency fluctuations of the laser-diode bias current. To quantitate this effect, the sensitivity of the RF phase noise to laser bias-current fluctuations was calculated by adding a small 5-Hz current modulation to the 31-mA direct-current bias [6]. This current modulation induced a phase-noise sideband at 5 Hz from the 100 MHz carrier, from which a bias current-to-phase noise conversion constant was derived.

The bias-current source noise was then measured from 1 Hz to 100 kHz, and the expected level of phase noise due to the bias-current fluctuations was calculated. This value is compared to the actual measured phase noise of the DFB laser-diode fiber-optic link in Fig. 3. The measured phase noise of the fiber-optic link is 35 dB above the theoretical minimum level set by the bias-current source noise. Therefore, it was concluded that bias-current source fluctuations are not responsible for the residual phase noise of the DFB laser link.

An investigation of the sensitivity of the phase noise and SNR to both optical reflections and temperature excursions was performed for the DFB laser link. Figure 4 illustrates the apparatus used for measuring the phase noise and SNR versus back-reflected power and laser-chip temperature. The laser was operated at a constant current

of 31 mA for an optical power output level of 1.7 mW, 2.3 dBm (dBm = dB with respect to 1 mW of optical power). The temperature of the thermoelectric cooler inside the isolated laser package was maintained to within 0.01°C, adjustable in 0.1° steps [7].

Figure 5 illustrates the results of measurements of the RF phase noise at 100 MHz as a function of back-reflected power for the DFB link. The noise appears to depend linearly on the amount of back-reflected optical power. The top curve in Fig. 5 represents the effect of a reflection of -18.5 dBm (return loss = 20.8 dB). These results suggest a simple relation which describes the amount of isolation required to render the laser insensitive to a reflection of a given power level,  $P_R$ ,

$$\begin{aligned} I(P_R) &= S_\phi(1 \text{ Hz}, P_R) - S_\phi(1 \text{ Hz}, 0) \\ &= -65 - (-115) \\ &= 50 \text{ dB} \end{aligned} \quad (1)$$

where  $I(P_R)$  is the additional isolation (in excess of the 30-dB internal isolator) in dB,  $S_\phi(1 \text{ Hz}, P_R)$  is the phase noise measured at the maximum reflected power level of  $P_R = -18.5 \text{ dBm}$  from Fig. 5, and  $S_\phi(1 \text{ Hz}, 0)$  is the phase noise measured with no reflected power (return loss =  $\infty$ ) from Fig. 3. Therefore, to achieve the optimum phase-noise performance at 1 Hz with a reflection of  $P_R$ , it appears that 50 dB of additional isolation is required in addition to the 30-dB internal isolator, for a total of 80 dB.

The effect of back-reflected power on the high-frequency-intensity noise of the laser was investigated by looking at the wideband noise on a spectrum analyzer display while changing the amount of back-reflected optical power. However, the laser-intensity noise did not change linearly over the entire range of back-reflected power. For the same 30-dB increase in reflected power from  $P_R = -45 \text{ dBm}$  to  $-15 \text{ dBm}$ , only a 10-dB increase in the SNR was observed. The low-frequency roll-off of the optical receiver did not permit study of the intensity-noise at frequencies below 10 MHz, but the laser-intensity noise appeared to increase as  $1/f$ .

Using the results of the RIN measurements to determine the isolation required to render the laser insensitive to optical reflections, the following is obtained:

$$\begin{aligned} I(P_R) &= N_{RIN}(100 \text{ MHz}, P_R) - N_{RIN}(100 \text{ MHz}, 0) \\ &= 10 \text{ dB} \end{aligned} \quad (2)$$

where  $N_{RIN}(100 \text{ MHz}, P_R)$  is the RF noise-power level measured with the maximum optical reflection, and  $N_{RIN}(100 \text{ MHz})$  is the RF noise-power level with minimum optical reflection.

The value  $I(P_R)$  obtained from the SNR measurement at 100 MHz is very different from the value obtained from the phase-noise measurement at 1 Hz from the carrier. Since the laser-intensity noise increases as  $1/f$  at lower frequencies, a simple measurement of the intensity noise versus reflected power at the RF modulation frequency leads to a grossly incorrect estimate of the amount of isolation required for good close-to-carrier phase-noise performance.

The phase-noise level and SNR were found to be extremely sensitive to variations of the laser chip temperature, as illustrated in Figs. 6 and 7. Both the phase noise and RIN showed a periodic dependence on the chip temperature with a 3- to 5-dB amplitude over a period of  $1^\circ\text{C}$ . Figure 8 vividly illustrates the periodic variation of the relative group delay for the 4-km link as the laser-diode temperature was swept slowly from  $25^\circ$  to  $50^\circ\text{C}$ . The trend of the data follows the expected parabolic group delay-versus-wavelength characteristic for single-mode fiber with linear dispersion, but the rapid periodic fluctuation in the group delay is unexpected and, at present, unexplained.

### III. Relationship of Laser-Intensity Noise to RF Phase Noise

In [2], it is shown that the intensity fluctuations of the laser will appear as close-to-carrier phase noise on an RF modulation signal. However, the more fundamental problem concerns the cause of the increased laser-intensity noise at low frequencies. Present theoretical models for semiconductor laser noise [8,9] predict laser-intensity noise to be white at low frequencies. However, a  $1/f$  dependence of the low-frequency intensity noise has been observed in the optical output of DFB lasers [10,11], in agreement with the observations.

If the low-frequency laser-intensity noise were white, as predicted by theory, the close-to-carrier phase noise of fiber-optic links would also be white, and would have the same value as the SNR at larger offset frequencies. At present, however, it appears that the unexplained  $1/f$  character of the laser-intensity noise limits the close-to-carrier phase-noise performance of analog fiber-optic links. Therefore, further research efforts to understand and reduce this noise are imperative for improved close-to-carrier phase-noise performance of fiber-optic links.

Other research [10] corroborates the measurements of this study, which indicate that optical reflections cause an increase in the magnitude of the  $1/f$  component of the laser-intensity noise. Therefore, it would seem that any reflections in the coupling optics, insufficient antireflection (AR) coating on the collimating lens, or reflections from the optical-isolator components will cause increased low-frequency laser-intensity noise. Thus, the residual level of close-to-carrier RF phase noise from Fig. 2 measured with minimum reflected power may be caused by reflections from the optical surfaces between the laser and the optical isolator. This may explain why lasers with different magnitudes of RIN can display equal phase-noise magnitudes when no excess reflections are induced. More study of the effect of residual reflections from coupling optics is required to quantify their contribution to low-frequency intensity noise.

Finally, it was observed that the conversion of low-frequency intensity noise to close-to-carrier RF phase noise may possibly be used to advantage to stabilize the output of laser diodes. Since laser intensity is coupled to laser frequency and phase through the carrier-density of the gain medium, the phase noise induced on an RF modulation signal represents a "probe" of the laser carrier density fluctuations. Consequently, it may be possible to use a continuous RF phase-noise measurement to generate an error signal to feed back on the laser bias current, thereby stabilizing the laser intensity, frequency, and phase fluctuations. These ideas are the subjects of ongoing research by the authors.

### IV. Design Considerations for Phase-Stable Semiconductor Laser Transmitters

To obtain optimum phase-noise performance from fiber-optic links subjected to unpredictable amounts of optical reflection, a fiber-optic transmitter package should incorporate the elements outlined in the preceding sections: high optical isolation, low-reflection coupling optics, and excellent temperature control. Conceptually, a transmitter employing these elements is illustrated in Fig. 9. It employs a wide-bandwidth, low-noise laser chip, a molded-glass aspheric collimating lens (AR-coated and located close to the laser-diode output facet), and two optical isolators in series to provide at least 60 dB of isolation. Even this amount of isolation may not be adequate for some applications where extremely high levels of reflected power are present. In these cases, additional external isolation could be added, provided the return loss of the external isolator is adequate and the intervening fiber is not subjected to flexure.

The key features of this design are that the laser beam is collimated and expanded using a single lens, and that the only back-reflection into the laser occurs within a few millimeters of the laser facet. It has been found [12] that optical reflections which occur within 5 mm of the laser-diode facet degrade noise performance very little. Then, the reflection from the first surface of the 60-dB isolator can be practically eliminated by an AR coating and a slight off-axis tilt. Following the isolator, a second molded lens would then focus the beam into a single-mode fiber.

The key component for the realization of this design is the molded glass aspheric lens, which provides the highly collimated beam necessary for maximum isolator performance. These types of lenses are currently used to focus the laser outputs in compact disc players, and represent a mature technology; however, the beam parameters of the laser must be tightly controlled in processing to insure efficient power transfer through the lens. An added benefit of using multiple isolators is reduced sensitivity of the isolation to fluctuations in ambient temperature.

## V. Summary and Conclusions

In the study described here, the close-to-carrier phase noise added to RF modulation signals by fiber-optic links based on direct current modulation of a semiconductor laser diode was characterized. This additive RF phase noise is due to the low-frequency laser-diode intensity noise, and can be significantly larger than the signal-to-noise ratio of the link at the RF modulation frequency. Thus, the recent improvements in semiconductor laser-

noise performance have not necessarily resulted in fiber-optic links with improved RF phase-noise performance.

It has been determined that the magnitude of the RF phase noise due to the laser-intensity noise is sensitive to the amount of optical power reflected into the laser cavity. Also, the phase-noise magnitude, SNR, and relative group delay of fiber-optic transmitters have been found to be extremely sensitive to laser-chip temperature. At present, it appears that the intensity noise generated in the semiconductor laser diode is the principal limitation of close-to-carrier phase stability in analog fiber-optic systems, but this point requires further study.

The results of the study suggest several key features which appear to be prerequisites for achieving improved RF phase stability in a fiber-optic transmitter. The key features are

- (1) greater than 60-dB optical isolation of the laser diode from all reflections
- (2) less than 5 mm of distance from the laser-diode facet to the first optical reflection
- (3) excellent temperature control of the laser chip, to better than 0.1°C

Such a design should permit phase-stable RF transmission over fiber-optic links subjected to time-varying reflections due to vibration and flexure of the fiber-optic cable. Work is in progress by the authors to characterize the performance of fiber-optic systems which incorporate these features to various degrees.

## Acknowledgment

The work described in this article was performed by the Jet Propulsion Laboratory, California Institute of Technology, under contract to the United States Air Force Rome Air Development Center.

## References

- [1] R. T. Wang, G. J. Dick, and D. M. Strayer, "Operational Parameters for the Superconducting Cavity Maser," *Proc. 12th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Tyson's Corner, Virginia, pp. 345-354, November 29-December 1, 1989.
- [2] K. Y. Lau, "Microwave phase stability of directly modulated semiconductor injection lasers," *Appl. Phys. Lett.*, vol. 52, no. 17, pp. 1377-1378, April 25, 1988.
- [3] Hewlett-Packard, "Understanding and Measuring Phase Noise in the Frequency Domain," Application Note 207, Palo Alto, California, October 1976.
- [4] Ortel Corporation, "3510B RF/Microwave Fiber-Optic Transmitter," manufacturer's data sheet, Alhambra, California, February 1988.
- [5] Mitsubishi Electric Corporation, "FU-45SDF-3 DFB-Laser Diode Module," manufacturer's data sheet, Tokyo, Japan, 1989.
- [6] ILX Lightwave Corporation, "Model LDX-3620 Ultra-Low Noise Current Source," manufacturer's data sheet, Bozeman, Montana, 1989.
- [7] ILX Lightwave Corporation, "Model LDT-5910 Thermoelectric Temperature Controller," manufacturer's data sheet, Bozeman, Montana, 1989.
- [8] K. Vahala and A. Yariv, "Semiclassical Theory of Noise in Semiconductor Lasers: Parts I and II," *IEEE J. Quantum Electron.*, vol. QE-19, no. 6, pp. 1096-1109, June 1983.
- [9] P. Spano, S. Piazzolla, and M. Tamburini, "Phase Noise in Semiconductor Lasers: A Theoretical Approach," *IEEE J. Quantum Electron.*, vol. QE-19, no. 7, pp. 1195-1199, July 1983.
- [10] Y. C. Chen, "Phase noise characteristics of single mode semiconductor lasers with optical feedback," *Appl. Phys. Lett.*, vol. 44, no. 1, pp. 10-12, January 1, 1984.
- [11] K. Kikuchi, "Effect of  $1/f$ -Type FM Noise on Semiconductor Laser Linewidth Residual in High-Power Limit," *IEEE J. Quantum Electron.*, vol. 25, no. 4, pp. 684-688, April 1989.
- [12] K. Sato, "Intensity Noise of Semiconductor Laser Diodes in Fiber-Optic Analog Video Transmission," *IEEE J. Quantum Electron.*, vol. QE-19, no. 9, pp. 1380-1391, September 1983.

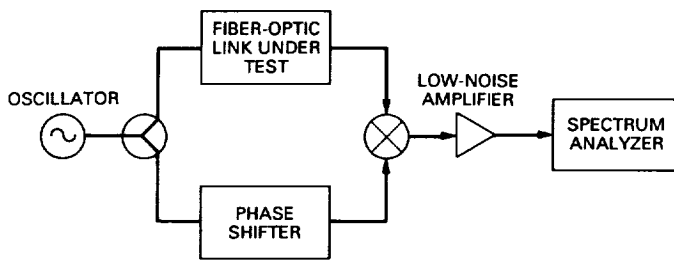


Fig. 1. RF phase-noise measurement system.

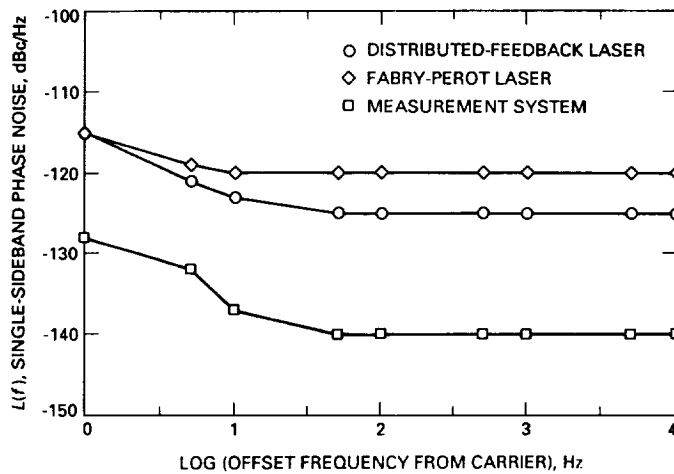


Fig. 2. RF phase noise at 100 MHz for Fabry-Perot and DFB laser-diode transmitters; measurement-system noise.

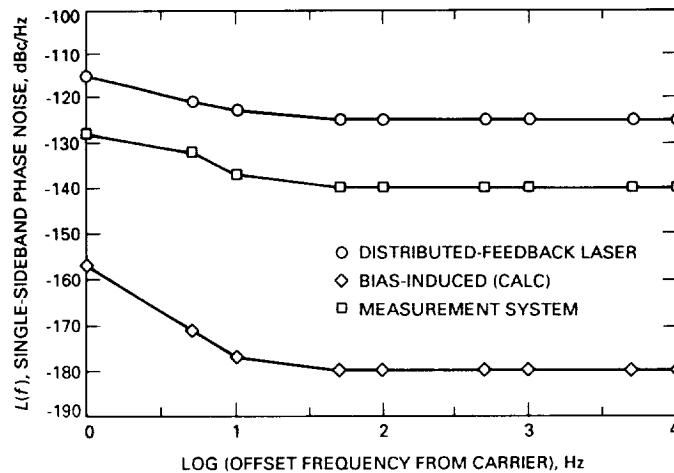


Fig. 3. RF phase noise for DFB laser transmitter, measurement noise floor, and calculated current-source-induced phase noise.

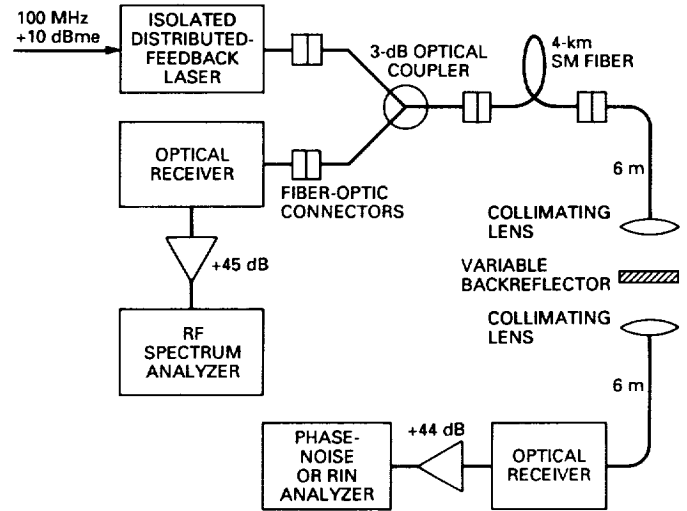


Fig. 4. Measurement system for investigation of RF phase noise and laser intensity versus back-reflected optical power.

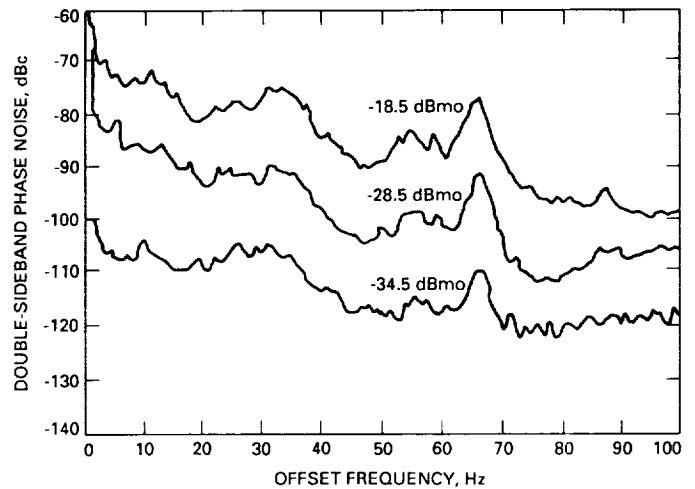


Fig. 5. RF phase noise versus reflected optical power for DFB transmitter with 30-dB internal isolator.

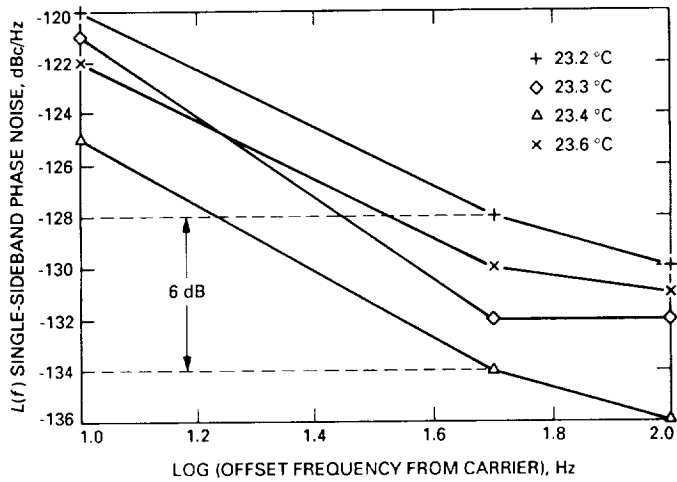


Fig. 6. RF phase noise of DFB laser transmitter versus laser-chip temperature.

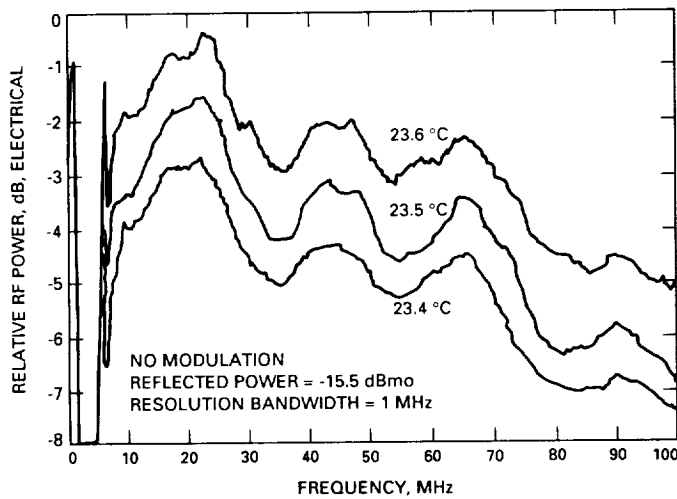


Fig. 7. Intensity noise of DFB laser transmitter versus laser-chip temperature.

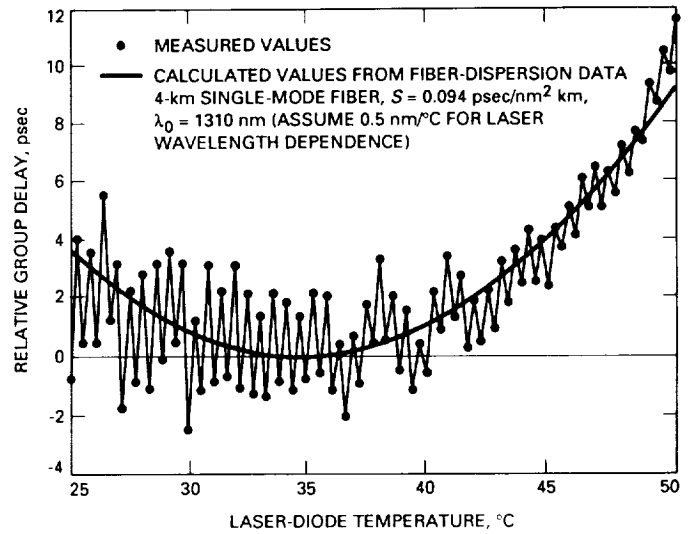


Fig. 8. Relative group delay of DFB laser transmitter versus laser-chip temperature.

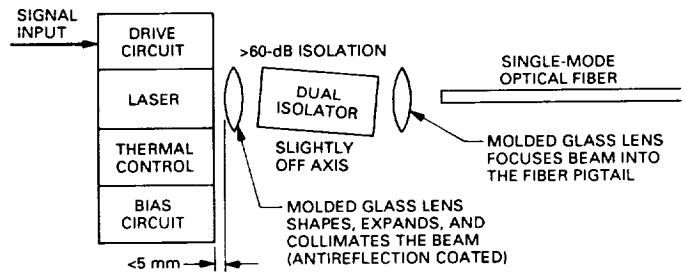


Fig. 9. Proposed laser transmitter design concept for improved phase stability.

# Navigational Utility of High-Precision Radio Interferometry for Galileo's Approach to Jupiter

W. M. Folkner

Tracking Systems and Applications Section

*The effect of very long baseline interferometry (VLBI) measurements of 2-nanoradian (nrad) accuracy has been studied for use in Galileo's approach to Jupiter's moon Io. Of particular interest is reducing the error in the minimum altitude above Io's surface. The nominal tracking strategy includes Doppler, range, and onboard optical data, in addition to VLBI data with 25-nrad accuracy. For nominal data, the altitude error is approximately 250 km with a data cutoff of 19 days before closest approach to Io. A limited number (two to four) of 2-nrad VLBI measurements, simulating a demonstration of improved VLBI data, were found to reduce the altitude error by 10 to 40 percent. Improving the accuracy of the VLBI measurements of the nominal tracking strategy to 2 nrad, to simulate the results from an operational few-nrad VLBI capability, was found to reduce the altitude error by an approximate factor of four. This reduction in altitude error is attributed to the ability that VLBI data give to help determine the along-track component of Jupiter's ephemeris. This capability complements the ability of the onboard optical data to determine the radial and cross-track components of Jupiter's ephemeris.*

## I. Introduction

Angular tracking through very long baseline interferometry (VLBI) has been routinely performed for the Voyager spacecraft [1] and is planned for the Galileo spacecraft. Using a technique called delta differential one-way ranging ( $\Delta$ DOR), the geometric delay of a radio signal from the spacecraft between two tracking stations is differenced with the delay of the signal from a nearby natural radio source. This difference provides a measurement of the angle between the spacecraft and the radio source. A  $\Delta$ DOR measurement along one baseline determines one component of the angular plane-of-sky spacecraft position.

Measurements on two nearly orthogonal baselines can provide complete plane-of-sky position information. The accuracy of the  $\Delta$ DOR measurement is about 50 nrad for the Voyager spacecraft and is expected to be about 25 nrad for the Galileo spacecraft using the present operational VLBI system.<sup>1</sup> An angular accuracy of 25 nrad corresponds to about 20 km over the Earth-Jupiter distance.

<sup>1</sup>J. B. Thomas, "An Error Analysis for Galileo Angular Position Measurements with the Block I  $\Delta$ DOR System," JPL Engineering Memorandum No. 335-26 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 11, 1981.

Improved VLBI technology promises to provide more accurate angular-position measurements. Specifically, the local-reference-frame technique [2] uses scans of the spacecraft and multiple radio sources to cancel many error sources. The resulting angular accuracy could reach the 2-nrad level. This order-of-magnitude improvement in VLBI measurement capability will not necessarily result in a similar orbit determination improvement. Other data, such as onboard optical data, may provide information that makes the VLBI data redundant. Random forces, due to solar pressure or gas leaks, may limit the utility of the improved VLBI data.

In order to study quantitatively the effect of high-precision VLBI measurements on orbit determination, a covariance study was performed on the approach of the Galileo spacecraft to Jupiter. This case was chosen since earlier analyses<sup>2</sup> [3] indicated that the VLBI data could improve orbit determination by reducing uncertainty about the Jovian ephemeris. This is accomplished by detecting the planet's gravitational signature on the motion of the spacecraft. The present study, while specific, extends the previous analyses by utilizing the full Orbit Determination Program (ODP), including random-force models and onboard optical data.

Starting from the previous work of Pollmeier,<sup>3</sup> covariance analyses were performed to examine how high-precision  $\Delta$ DOR measurements might affect the expected orbit-determination accuracy. For this study, the high-precision  $\Delta$ DOR measurements were treated as regular  $\Delta$ DOR measurements with improved accuracy. Three groups of cases are presented here. The first group examines the effect of a limited number (two to four measurements per baseline) of high-precision  $\Delta$ DOR measurements. These are in addition to the nominal radio metric and optical data on the standard Galileo orbit determination. These few measurements could be included in an orbit analysis to provide a demonstration of the local-reference-frame technique. They also could be included in a post-encounter analysis to evaluate the accuracy of the local-reference-frame technique for spacecraft navigation. One case considered by Pollmeier includes increased Doppler accuracy, which would be attained using

an X-band (8.1-GHz) uplink instead of the nominal S-band (2.3-GHz) uplink, for improved orbit determination. The second group of cases included in this study evaluates the effect of adding a limited number of high-precision  $\Delta$ DOR measurements to this improved Doppler data. For the final cases, all normally scheduled  $\Delta$ DOR measurements are treated as high-precision measurements, as though the local-reference-frame technique was an operational capability. While this is not likely to be true of Galileo's Jupiter approach, it serves to show the potential of improved VLBI-tracking data.

## II. Trajectory and Mapping

These analyses are based on the trajectory for the Io encounter, scheduled to take place on December 7, 1995.<sup>4</sup> Figure 1 is a plot of the approach trajectory in the ecliptic plane as seen by an inertial observer. Jupiter is traveling from left to right catching up to the spacecraft. The points are shown at four-hour intervals. Figure 2 is a plot of the approach trajectory in the ecliptic plane centered on Jupiter with points at one-hour intervals. Galileo makes a close encounter with Jupiter's moon Io approximately five hours before closest approach to Jupiter. The Io encounter is used to provide a gravitational assist in establishing the Jupiter-centered orbit. A major navigation requirement is that Galileo remain 500 km or higher above the surface of Io to avoid possible collision with volcanic ejecta. The trajectory design has Galileo pass 1000 km above Io. Consultations with the Galileo orbit determination team of F. T. Nicholson, D. W. Murrow, and V. M. Pollmeier suggest that there are two times when additional accuracy will be desired, namely at the 26-hour data cutoffs preceding trajectory correction maneuvers TCM-27 and -28. These cutoffs occur at Io-19 days 2 hours and at Io-11 days 2 hours.

For local-reference-frame measurements, a number of quasars are needed about the spacecraft. Figure 3 is a plot of the spacecraft trajectory in the plane of the sky with points at four-day intervals. The low declination of the approach trajectory will cause observations at the Goldstone and Madrid sites to be at low elevation. This will restrict the choice of quasars for the local reference frame and may reduce the accuracy of local-reference-frame  $\Delta$ DOR measurements. Three of the radio sources shown in Fig. 3

<sup>2</sup>R. N. Treuhaft and J. S. Ulvestad, "Using Gravitational Signatures for Target-Relative Angular Tracking During Planetary Approach," JPL Interoffice Memorandum 335.3-88-76 (internal document), Jet Propulsion Laboratory, Pasadena, California, July 11, 1988.

<sup>3</sup>V. M. Pollmeier, "Io Delivery Uncertainties and Parametric Variation Studies for the Jupiter Approach of the October 10 Gaspra Ida Trajectory," JPL Interoffice Memorandum GLL-NAV-89-31 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 16, 1989.

<sup>4</sup>J. R. Johannessen, "Reference Integrated Interplanetary, Probe, and Orbiter Trajectories for the 1989 VEEGA with Gaspra and Ida Asteroid Flyby," JPL Interoffice Memorandum 312-88.4-1461 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 25, 1988.

are from the Magellan catalog<sup>5</sup> with the remainder from a VLBI survey [4].

The initial spacecraft state was specified at Io encounter minus 142 days, which is after the separation of the probe from the orbiter and after the orbit-deflection maneuver (ODM). Maneuvers included in this analysis are the ODM-cleanup maneuver (TCM-26) at encounter minus 100 days and trajectory-correction maneuvers, TCM-27 and -28. The orbit-determination results are mapped to the Io-centered B-plane coordinate system<sup>6</sup> referred to the Jupiter mean equator of 1950 at time of closest approach to Io. Since the spacecraft trajectory lies mainly in Jupiter's equatorial plane and is not greatly deflected by Io, the B-plane component  $B \cdot T$  corresponds within a few percent to the distance from the center of Io. The  $B \cdot T$  error corresponds to the error in altitude above Io, which is the error of interest.

### III. Error Modeling

The error modeling used here is adopted from Pollmeier.<sup>7</sup> In the covariance analysis, many parameters are estimated along with the spacecraft state while other parameters are considered. In a consider analysis [5], errors in estimated parameters due to (nonestimated) considered parameters are added (in quadrature) to the computed error (which resulted from data noise and a priori errors in the estimated parameters) to form the total error. An error from a considered parameter which forms a large portion of the total error may indicate a need for a different error modeling for that parameter. The estimated and considered parameters for the standard cases are listed in Table 1 along with a priori error information. The diagonal components of the covariance matrices for the Earth-Jupiter ephemeris,<sup>8</sup> the Galilean satellite

ephemerides,<sup>9</sup> and the station locations<sup>10</sup> are listed in Table 1. Since the ephemerides are estimated and only one quasar is used for  $\Delta$ DOR measurements, the quasar can be used to define the orientation of the reference frame. If more than one quasar were used, a relative quasar location error would have to be included.

The standard data types are S-band two-way Doppler and range, X-band VLBI data, and onboard optical data. The nominal data schedule is outlined in Table 2. The minimum elevation for radio metric data is 15 deg for range and Doppler and 7 deg for  $\Delta$ DOR. Radio metric data are taken more often just after the probe release (Io-142 to Io-136), around the ODM-cleanup maneuver at Io-100 days, and from Io-60 days to encounter.

For the purpose of the local-reference-frame  $\Delta$ DOR demonstration, an additional set of measurements was scheduled with two points for each of the Goldstone-Canberra and Goldstone-Madrid baselines in the week preceding TCM-27 and two more per baseline between TCM-27 and -28. The times, shown in Table 3, were chosen to allow about two days of processing time prior to the data cutoffs. When included, these measurements were independently weighted and merged with the rest of the data set. The high-precision  $\Delta$ DOR measurements were weighted at either 3 cm or 1 cm (corresponding to a 100-psec or 33-psec delay error). The plane-of-sky angular error depends on the projected baseline length. For example, a 1-cm  $\Delta$ DOR error and a 5000-km projected baseline correspond to 2-nrad angular accuracy.

Station-location and media errors, which are treated as consider parameters for  $\Delta$ DOR measurements, are estimated in the local-reference-frame technique. For cases which included only a limited number of high-precision  $\Delta$ DOR measurements, the station-location and media errors did not significantly contribute to the total orbit-determination error. But when all  $\Delta$ DOR measurements were treated as high precision, analyses were done both with the standard considered-error models (with station-location and media errors for the  $\Delta$ DOR measurements considered) and with modified error modeling (with station-location and media errors for  $\Delta$ DOR measurements removed).

<sup>5</sup> J. S. Ulvestad and O. J. Sovers, "Preliminary VLBI catalog for Magellan," JPL Interoffice Memorandum 335.3-89-14 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 30, 1989.

<sup>6</sup> For a description of the B-plane coordinate system, see G. W. Spier, "Design and Implementation of Models for the Double Precision Trajectory Program (DPTRAJ)," JPL Technical Memorandum 33-451 (internal document), Jet Propulsion Laboratory, Pasadena, California, pp. 103-116, 1971.

<sup>7</sup> V. M. Pollmeier, *op. cit.*

<sup>8</sup> E. M. Standish, "The Covariance of Venus, Jupiter, and the Earth," JPL Interoffice Memorandum 314.6-969 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 1, 1988.

<sup>9</sup> D. W. Murrow, "A Covariance for the Galilean Satellites for ODP Analysis," JPL Interoffice Memorandum 314.3-771 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 5, 1987.

<sup>10</sup> D. W. Murrow and F. T. Nicholson, "Station Location Covariance," JPL Interoffice Memorandum GLL-NAV-88-50 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 2, 1988.

## IV. Standard Case Results

Table 4 lists three sets of orbit-determination results at the Io-19- and -11-day cutoffs. All three sets include the standard error models and the nominal data outlined in the previous section. The set labeled Nominal includes 16  $\Delta$ DOR points per baseline for the Io-19-day cutoff and 18  $\Delta$ DOR points per baseline for the Io-11-day cutoff, but does not include high-precision  $\Delta$ DOR data. The Demo-3-cm set includes the demonstration measurements from Table 3 weighted at 3 cm. The Demo-1-cm set includes the same demonstration measurements weighted at 1 cm.

The errors are tabulated in the Io-centered B-plane coordinate system for completeness. The impact parameter  $\vec{B}$  has components  $B \cdot T$ , in the plane of the trajectory, and  $B \cdot R$ , perpendicular to the trajectory plane. Of particular interest is the error in altitude above Io, which corresponds to the column labeled  $B \cdot T$ . The linearized time of flight (LTOF) parameterizes the position along the trajectory. The LTOF error (in seconds) can be multiplied by the velocity at infinity,<sup>11</sup> 14.9 km/sec, to compare it to the impact-parameter errors  $B \cdot R$  and  $B \cdot T$ . The LTOF error is thus the largest error by about 15 percent. Inclusion of more or better  $\Delta$ DOR data will result in a comparable improvement in  $B \cdot T$  and LTOF. That is, a reduction of 20 percent in the LTOF error is accompanied by a corresponding 20 percent reduction in the  $B \cdot T$  error. This is because the dominant error is the Jupiter along-track error which projects by a fixed amount in the  $B \cdot T$  and LTOF directions. The  $B \cdot R$ , which is perpendicular to the along-track error, is not affected much by improvements in the  $\Delta$ DOR data since the onboard optical data provide better information in the cross-track direction.

The important error is the error in altitude above Io ( $B \cdot T$  error) at the Io-19- and -11-day cutoffs. The Demo-3-cm set shows an altitude error improvement of 19 percent over the Nominal set at the Io-19-day cutoff and a 21 percent improvement at Io-11 days. The Demo-1-cm set shows an improvement of 36 percent over the Nominal set at Io-19 days and 38 percent at Io-11 days.

## V. Demonstration of High-Precision $\Delta$ DOR with Improved Doppler Data

Since improved orbit-determination accuracy is desired over standard results, a case with improved Doppler data accuracy is being considered.<sup>12</sup> The improved Doppler accuracy would be achieved through use of an X-band uplink

and downlink instead of the nominal S-band frequency. The cases presented here examine whether a demonstration of the high-precision  $\Delta$ DOR system could yield a significant improvement in orbit-determination accuracy in the presence of improved Doppler data.

For these cases, Doppler accuracy was improved from 1 mm/sec to 0.2 mm/sec due to reduced instrumental errors for the higher-frequency link. The ionosphere errors were reduced by a factor of ten (from 75 to 7.5 cm) due to the  $1/f^2$  behavior of the ionosphere delay. The improved Doppler accuracy increased the sensitivity of the orbit determination to station-location uncertainty. Rather than include a large considered station-location error, these cases used the improved Doppler data to estimate station locations. The data schedule was the same as for the standard cases.

Table 5 lists the results for the improved Doppler cases. The altitude error for Demo 3 cm is 14 percent lower than the Nominal case at Io-19 days and 11 percent lower at Io-11 days. The Demo-1-cm results are 29 percent lower than the Nominal at Io-19 days and 30 percent lower at Io-11 days.

## VI. High-Precision $\Delta$ DOR as an Operational Data Type

If high-precision  $\Delta$ DOR is demonstrated to be a viable navigational tool, it will be possible to schedule such measurements regularly during the approach phase of a mission. For the cases presented in Table 6, the  $\Delta$ DOR measurements from Table 2, comprising VLBI measurements scheduled for mission navigation, were included with 1-cm weight. Demonstration measurements from Table 3 were not included. The weights for other data types were the same as in the standard case. The All- $\Delta$ DOR-1-cm set with standard models retains the same considered-error models for media, station locations, and quasar location as the regular  $\Delta$ DOR measurements. The computed  $B \cdot T$  error at the Io-19-day cutoff shows improvement by a factor of three over the standard-case result with Nominal data. The contribution of media errors for this case is nearly as large as the computed error. For local-frame measurements, media effects can be estimated separately and included in the measurement error. For the All- $\Delta$ DOR-1-cm case with modified models, the ODP was modified to ignore the considered media and station-location errors for the  $\Delta$ DOR measurements while retaining those considered errors for the range and Doppler measurements. In this case, the considered errors are reduced and the total  $B \cdot T$  error is smaller than the standard nominal case error by nearly a factor of four.

<sup>11</sup> Spier, op. cit.

<sup>12</sup> Pollmeier, op. cit.

Figure 4 shows the projected altitude error at Io versus data cutoff time for three specific data sets. The curve labeled Nominal is based on the standard S-band data models. The filled squares at the bottom of the plot denote days when pairs of regular  $\Delta$ DOR measurements are scheduled, one measurement per baseline. The orbit analysis includes seven pairs of  $\Delta$ DOR measurements prior to Io-50 days. The curve labeled Demo 1 cm includes the extra  $\Delta$  DOR measurements listed in Table 3 (denoted by filled triangles in Fig. 4). The projected altitude error is reduced after the Io-50 days. The curve labeled Demo 1 cm includes the extra  $\Delta$  DOR measurements listed in Table 3 (denoted by filled triangles in Fig. 4). The projected altitude error is reduced after the demonstration measurements until a few days before Io encounter, when the onboard optical data become dominant. The curve with All  $\Delta$ DOR 1 cm is for the case where the regularly scheduled measurements are treated as high-precision measurements. The projected altitude error is reduced from the Nominal curve until the final few days.

## VII. Effect of the Jupiter Ephemeris Error

Inclusion of the Earth-Jupiter and Galilean satellite ephemerides in the estimation list makes it difficult to see what improvement is taking place with changes in data. Analyses were performed with some of the estimated parameters removed to discover which estimated error sources were most important. Table 7 lists the computed errors when one of the following was removed from the estimation list: the Earth-Jupiter ephemeris, the radial stochastic spacecraft acceleration, or the Galilean satellite ephemerides. The data sets used were from the Nominal S-band, Demo-1-cm S-band, and All- $\Delta$ DOR-1-cm cases. The considered errors are not listed since changing the estimation list also changes the computed error and the effect of considered parameters.

In Table 7, the standard set estimates all parameters listed in Table 1. The next line lists the computed errors with the Galilean satellite ephemerides parameters removed from the estimation list. This produced little change in the  $B \cdot T$  and LTOF errors, but the  $B \cdot R$  error is reduced by a factor of two, although this error is not significant. The third line differs by the removal of the stochastic radial acceleration from the estimation list. However, it does include the satellite ephemerides. Dropping the stochastic acceleration improves the computed

error by 20 to 30 percent depending on the data. But the stochastic radial acceleration is not the dominant error and does not prevent improved  $\Delta$ DOR data from improving the orbit determination. The fourth and final line in Table 7 shows the effect of deleting the joint Earth-Jupiter ephemeris from the estimation list, while retaining the stochastic and satellite ephemeris. The computed errors are drastically reduced and improved  $\Delta$ DOR data have little effect on the solution.

Examination of the estimated Earth and Jupiter state errors shows that the main error is in the down-track Jupiter error. Figure 5 shows the estimated Jupiter down-track error at several times. The nominal data improve the down-track error by 35 percent at the Io-19-day cutoff. Improved  $\Delta$ DOR data significantly improve the down-track error but have almost no effect in the out-of-plane error. The estimated Earth state is better known than the Jupiter state and is not affected by the improved  $\Delta$ DOR data.

## VIII. Conclusion

This study has treated local-frame  $\Delta$ DOR measurements as regular  $\Delta$ DOR measurements with improved accuracy. It has shown that more accurate VLBI data improve orbit-determination accuracy. For the Galileo approach to Io, the errors in both the LTOF and  $B \cdot T$  components are reduced significantly, with the latter error corresponding to the error in altitude above Io. Two  $\Delta$ DOR measurements per baseline, with an accuracy of 1 cm, can produce an altitude accuracy improvement from 250 to 150 km at 19 days before Io encounter. A larger number of high-precision  $\Delta$ DOR measurements can produce a factor of four improvement in altitude determination. Thus, improved  $\Delta$ DOR measurements have the potential to greatly improve spacecraft navigation. Galileo's approach to Io provides a good opportunity to demonstrate that capability. The proper treatment of errors for high-precision VLBI measurements has yet to be determined. That includes determining which parameters are estimated as part of the measurement and which are included in the orbit-determination estimation. And there are still outstanding questions to be answered regarding reference-frame definition, nutation-series compatibility, and treatment of media and station errors, among other concerns.

## Acknowledgments

The author thanks F. T. Nicholson, V. M. Pollmeier, D. W. Murrow, M. H. Finger, R. N. Treuhaft, and J. M. Davidson for their input and cooperation during the course of this study.

## References

- [1] J. S. Border, F. F. Donovan, S. G. Finley, C. E. Hildebrand, B. Moultrie, and L. J. Skjerve, "Determining Spacecraft Angular Position With Delta VLBI: The Voyager Demonstration," AIAA/AAS Astrodynamics Conference, San Diego, California, August 9–11, 1982.
- [2] R. N. Treuhaft, "Deep Space Tracking in Local Reference Frame," *TDA Progress Report 42-94*, vol. April–June, Jet Propulsion Laboratory, Pasadena, California, pp. 1–15, July 15, 1988.
- [3] J. K. Miller and K. H. Rourke, "The Application of Differential VLBI to Planetary Approach Orbit Determination," *DSN Progress Report 42-40*, vol. May and June, Jet Propulsion Laboratory, Pasadena, California, pp. 84–90, May 1977.
- [4] R. A. Preston, D. D. Morabito, J. G. Williams, J. Faulkner, D. L. Jauncey, and G. D. Nicolson, "A VLBI Survey at 2.29 GHz," *Astronomical Journal*, vol. 90, pp. 1599–1641, 1985.
- [5] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Orlando, Florida: Academic Press, 1977.

**Table 1. Estimated and considered error terms**

Estimated parameters	
Parameter	a priori error
Spacecraft position	$10^8$ km
Spacecraft velocity	$10^8$ km/sec
Constant radial acceleration	$10^{-12}$ km/sec <sup>2</sup>
Stochastic radial acceleration	$10^{-12}$ km/sec <sup>2</sup> (a 5-day time constant)
Correction maneuvers	$10^{-5}$ km/sec each direction
Earth ephemeris	24 m radial 21 km down track 28 km cross track
Jupiter ephemeris	15 km radial 140 km down track 162 km cross track
Io ephemeris	5 km radial 36 km down track 22 km cross track
Considered parameters	
Parameter	a priori error
Station locations	75-cm spin radius 1-m longitude 10-m z-height 10-cm baselines
Radial solar pressure	0.171
Wet troposphere	4 cm
Dry troposphere	1 cm
Day ionosphere (S-band)	75 cm
Night ionosphere (S-band)	15 cm
Optical centerfinding errors	0.5 pixel, 0.5 line

**Table 2. Nominal data schedule for Galileo approach**

Data type	Weight	Time period	Schedule
Optical	0.5 pixel	Io-57 to Io	Approx. 1 per satellite per day
Range	1 km	Io-142 to Io-136	DSS 14,43 1 per pass
		Io-136 to Io-110	DSS 43 1 every other pass
		Io-110 to Io-98	DSS 14,43 1 per pass
		Io-97 to Io-60	DSS 43 1 every other pass
		Io-60 to Io	DSS 14,43 1 per pass
Doppler	1 mm/sec	Io-142 to Io-136	DSS 14,43,63 1 per hour per pass
		Io-136 to Io-110	DSS 43 1 per hour every other pass
		Io-110 to Io-98	DSS 14,43,63 1 per hour per pass
		Io-97 to Io-60	DSS 43 1 per hour every other pass
		Io-60 to Io	DSS 14,43,63 1 per hour per pass
$\Delta$ DOR	14 cm	Io-146 to Io-136	1 per baseline per week
		Io-136 to Io-60	1 per baseline per month
		Io-60 to Io	2 per baseline per week

**Table 3. Schedule of extra  $\Delta$ DOR measurements for local reference frame demonstration cases**

Baseline	Date	Time	Cutoff
14-43	12-Nov-1995	23:36	
14-63	13-Nov-1995	17:42	
14-63	16-Nov-1995	17:32	
14-43	16-Nov-1995	23:23	
	18-Nov-1995	15:45	Io-19d 2h
14-43	20-Nov-1995	23:10	
14-63	21-Nov-1995	17:16	
14-63	24-Nov-1995	17:07	
14-43	24-Nov-1995	22:57	
	26-Nov-1995	15:45	Io-11d 2h

**Table 4. Orbit-determination results for the Standard cases**

Error breakdown for Io-19-day 2-hour cutoff									
	Nominal			Demo 3-cm $\Delta$ DOR			Demo 1-cm $\Delta$ DOR		
	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec
Computed	20.0	248.4	18.9	19.3	199.0	15.2	19.3	157.9	12.1
Solar	0.4	12.6	1.0	0.3	9.5	0.7	0.2	0.5	0.5
Station	0.6	13.7	1.0	0.7	16.4	1.2	0.6	1.0	1.0
Media	0.9	50.5	3.8	1.1	52.9	4.0	1.3	37.5	2.8
Optical	3.1	6.8	0.7	3.2	11.6	1.1	3.3	12.8	1.2
Total	20.3	254.7	19.4	19.9	207.4	15.8	19.6	163.7	12.5
Error breakdown for Io-11-day 2-hour cutoff									
	Nominal			Demo 3-cm $\Delta$ DOR			Demo 1-cm $\Delta$ DOR		
	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec
Computed	16.1	119.5	9.2	15.4	93.8	7.2	14.9	71.9	5.5
Solar	0.1	1.3	0.1	0.1	1.6	0.1	0.0	1.7	0.1
Station	0.3	8.1	0.6	0.4	4.1	0.3	0.4	5.6	0.4
Media	0.7	15.7	1.2	1.1	1.5	1.5	1.1	22.0	1.7
Optical	3.8	46.7	3.8	3.8	35.5	3.0	3.9	26.5	2.3
Total	16.6	129.5	10.0	15.9	102.2	7.9	15.5	80.0	6.2

**Table 5. Orbit determination results for the Improved Doppler cases**

Error breakdown for Io-19-day 2-hour cutoff									
	Nominal			Demo 3-cm $\Delta$ DOR			Demo 1-cm $\Delta$ DOR		
	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec
Computed	19.7	219.0	16.7	19.4	183.4	14.0	19.2	150.7	11.4
Solar	0.2	8.1	0.6	0.2	7.1	0.5	0.2	5.8	0.4
Media	1.0	30.3	2.3	1.3	52.3	4.0	1.2	41.4	3.1
Optical	3.3	29.1	2.4	3.3	24.3	2.0	3.3	19.3	1.7
Total	20.0	223.1	17.0	19.7	192.4	14.7	19.5	157.6	12.0
Error breakdown for Io-11-day 2-hour cutoff									
	Nominal			Demo 3-cm $\Delta$ DOR			Demo 1-cm $\Delta$ DOR		
	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec
Computed	16.0	86.9	6.7	15.3	77.5	6.0	14.9	61.8	4.8
Solar	0.0	0.9	0.1	0.1	1.3	0.1	0.1	1.7	0.1
Media	0.6	4.9	0.4	0.9	11.3	0.9	0.9	15.5	1.2
Optical	3.7	42.7	3.5	3.7	35.0	2.9	3.8	23.5	2.0
Total	16.4	97.0	7.6	15.8	85.8	6.7	15.4	67.9	5.3

**Table 6. Orbit determination results for the All- $\Delta$ DOR-1-cm cases**

Error breakdown for Io-19-day 2-hour cutoff						
	Standard models			Modified models		
	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec
Computed	18.8	65.4	5.0	18.8	65.4	5.0
Solar	0.2	6.9	0.5	0.2	6.9	0.5
Station	0.4	14.2	1.1	0.0	1.3	0.1
Media	2.7	42.3	3.2	0.0	0.4	0.0
Optical	3.1	16.2	1.4	3.1	16.2	1.4
Total	19.3	81.2	6.2	19.1	67.8	5.2
Error breakdown for Io-11-day 2-hour cutoff						
	Standard models			Modified models		
	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec
Computed	15.0	41.9	3.2	15.0	41.9	3.2
Solar	0.1	5.4	0.4	0.1	5.4	0.4
Station	0.6	9.7	0.7	0.0	0.4	0.0
Media	1.2	36.3	2.8	0.0	0.4	0.0
Optical	3.9	15.7	1.4	3.9	15.7	1.4
Total	15.5	58.7	4.6	15.5	45.1	3.5

**Table 7. Computed errors with individual variables removed**

Computed errors for Io-19-day 2-hour cutoff									
	Nominal			Demo 1-cm $\Delta$ DOR			All $\Delta$ DOR 1 cm		
	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec
Standard set	20.0	248.4	18.9	19.3	157.9	12.1	18.8	65.4	5.0
No satellites	11.6	247.3	18.8	10.2	155.9	11.7	9.3	57.0	4.7
No stochastic	19.3	171.0	13.0	18.9	115.9	8.7	18.7	52.9	4.0
No Jupiter	18.3	44.7	3.3	17.9	38.0	2.8	17.9	37.6	2.8
Computed errors for Io-11-day 2-hour cutoff									
	Nominal			Demo 1-cm $\Delta$ DOR			All $\Delta$ DOR 1 cm		
	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec	$B \cdot R$ , km	$B \cdot T$ , km	LTOF, sec
Standard set	16.1	119.5	9.2	14.9	71.9	5.5	15.0	41.9	3.2
No satellites	8.8	113.2	8.7	6.0	67.1	5.3	5.9	34.0	2.8
No stochastic	15.9	61.8	4.6	14.9	39.1	2.9	14.9	33.4	2.4
No Jupiter	15.7	31.9	2.2	14.3	21.9	1.4	14.4	22.2	1.5

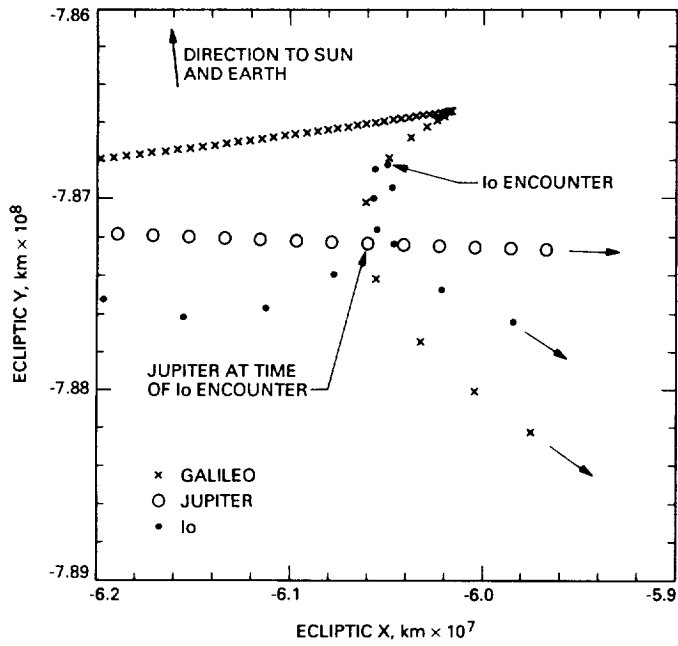


Fig. 1. Plot of the trajectories of Galileo, Jupiter, and Io in the ecliptic plane as seen by an inertial observer. Points are shown at four-hour intervals.

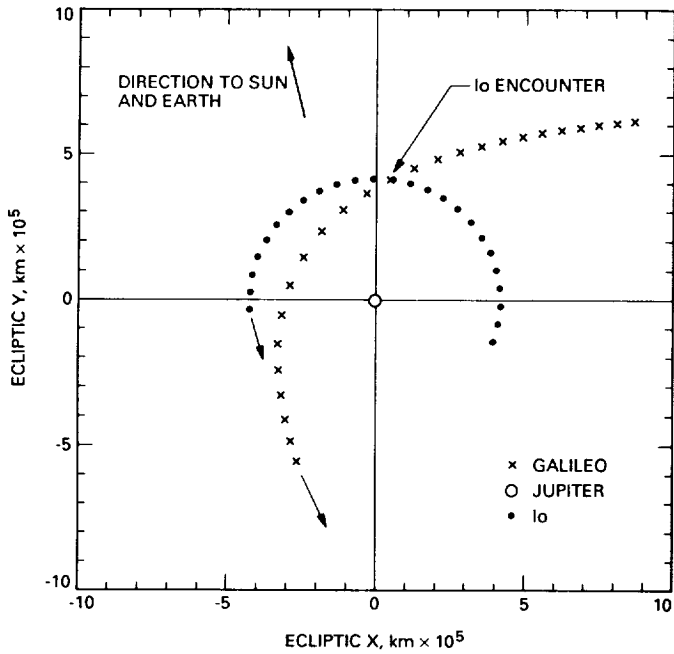


Fig. 2. Plot of the trajectories of Galileo, Jupiter, and Io in the ecliptic plane centered on Jupiter. Points are shown at one-hour intervals.

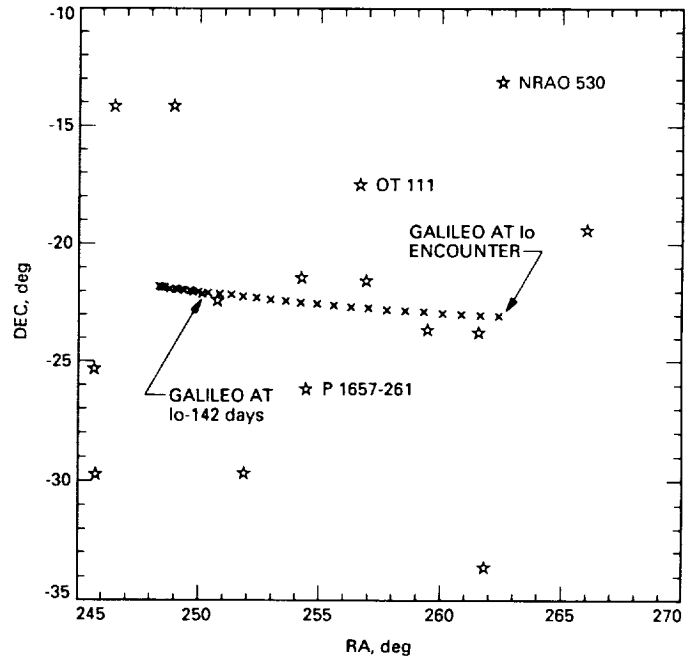


Fig. 3. Plot of the trajectory of Galileo in the plane of the sky. Points are shown at four-day intervals. The labeled radio sources are in the JPL Magellan catalog. The unlabeled sources are from the VLBI survey of [4].

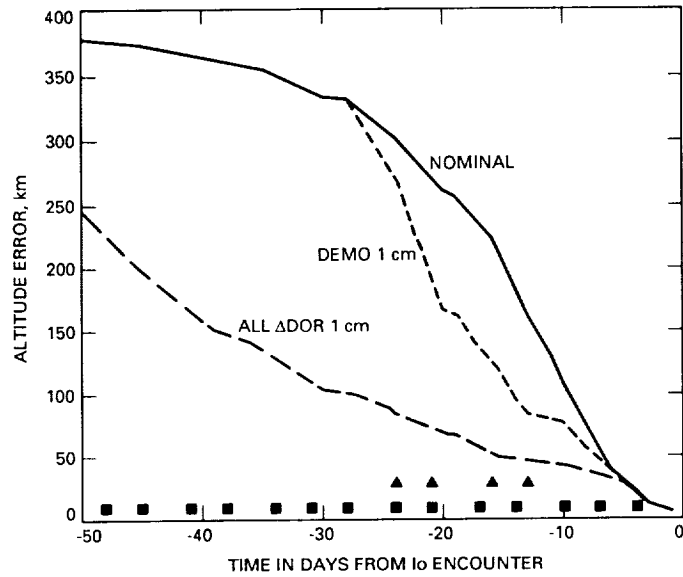


Fig. 4. Plot of the projected-altitude error at Io versus data-cutoff time for three data sets. The filled squares represent days when pairs of  $\Delta$ DOR measurements are scheduled. The triangles show the times of the demonstration-measurement pairs.

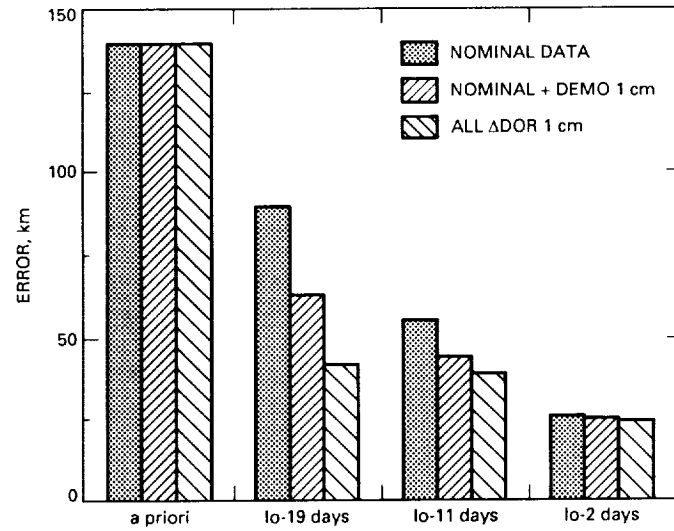


Fig. 5. The estimated Jupiter down-track error with different data sets.

38-61  
200343  
P-10  
August 15, 1990

N91-11978

# Linear-Phase Approximation in the Triangular Facet Near-Field Physical Optics Computer Program

W. A. Imbriale and R. E. Hodges  
Ground Antennas and Facilities Engineering Section

*Analyses of reflector antenna surfaces use a computer program based on a discrete approximation of the radiation integral. The calculation replaces the actual surface with a triangular facet representation; the physical optics current is assumed to be constant over each facet. This article describes a method of calculation using linear-phase approximation of the surface currents of parabolas, ellipses, and shaped subreflectors and compares results with a previous program that used a constant-phase approximation of the triangular facets. The results show that the linear-phase approximation is a significant improvement over the constant-phase approximation, and enables computation of  $100-1,000\lambda$  reflectors within a reasonable time on a Cray computer.*

## I. Introduction

One of the simplest reflector antenna computer programs is based on a discrete approximation of the radiation integral. This calculation replaces the actual reflector surface with a triangular facet representation so that the reflector resembles a geodesic dome. The physical optics (PO) current is assumed to be constant in magnitude and phase over each facet, so the radiation integral is reduced to a simple summation. This program has proven to be surprisingly robust and useful for the analysis of small reflectors, particularly when the near-field is desired and surface derivatives are not known.

It is natural to inquire whether a more sophisticated approximation of the PO surface current will yield more accurate results or permit the use of larger facets. In this article, a linear-phase approximation of the surface current is made. Within each triangular region, the resulting

integral is the two-dimensional Fourier transform of the projected triangle. This triangular-shape function integral can be computed in closed form [1]. The complete PO integral is then a summation of these transforms.

Once the current on the triangle is determined, the linear-phase approximation takes about three times longer to compute a field point than does the constant-phase approximation. Thus the time savings depend on reducing the number of triangles required to achieve convergence.

Examples are given for scattering from parabolas, ellipses, and shaped surfaces. The general trend is similar for all cases in that, depending upon the size of the triangles, there is an angular limit over which the solution is valid. This angular limit is significantly larger with the linear-phase approximation than with the constant-phase approximation. Thus the linear-phase approximation can be used to solve larger problems if core limitations are a

problem, or alternatively, a smaller number of triangles can be used to solve the same size problem.

## II. Analytical Details

The PO radiation integral over the reflector surface  $\Sigma$  can be expressed [2]

$$\mathbf{H}(\mathbf{r}) = -\frac{1}{4\pi} \int_{\Sigma} \left( jk + \frac{1}{R} \right) \hat{\mathbf{R}} \times \mathbf{J}_s(\mathbf{r}') \frac{e^{-jkR}}{R} ds' \quad (1)$$

in which  $\mathbf{r}$  designates the field point,  $\mathbf{r}'$  the source point,  $R = |\mathbf{r} - \mathbf{r}'|$  is the distance between them, and  $\hat{\mathbf{R}} = (\mathbf{r} - \mathbf{r}')/R$  is a unit vector. The PO surface current on the subreflector surface  $\mathbf{J}_s$  is expressed

$$\mathbf{J}_s(\mathbf{r}') = 2\hat{\mathbf{n}} \times \mathbf{H}_s(\mathbf{r}') \quad (2)$$

For the purpose of analysis, the true surface  $\Sigma$  is replaced by a contiguous set of  $N$ -plane triangular facets. These facets, denoted  $\Delta_i$ , are chosen to be roughly equal in size with their vertices on the surface  $\Sigma$ . Figure 1 shows a typical facet and its projection onto the  $x$ - $y$  plane. Let  $(x_i, y_i, z_i)$  represent the *centroid* of each triangle where the subscript  $i = 1, \dots, N$  is associated with a triangle. Then, the field obtained by replacing the true surface  $\Sigma$  by the triangular facet approximation is

$$\mathbf{H}(\mathbf{r}) = -\frac{1}{4\pi} \sum_{i=1}^N \int_{\Delta_i} \left( jk + \frac{1}{R} \right) \hat{\mathbf{R}} \times \mathbf{J}(\mathbf{r}') \frac{e^{-jkR}}{R} ds' \quad (3)$$

In Eq. (3),  $\mathbf{J}$  is now the equivalent surface current evaluated on the triangular facets. Since the triangles are small, it is expected that  $\hat{\mathbf{R}}$  and  $R$  do not vary appreciably over the area of a given facet. Thus, let  $\hat{\mathbf{R}}_i$  and  $R_i$  be the value obtained at the centroid  $(x_i, y_i, z_i)$  of each facet and approximate Eq. (3) by

$$\mathbf{H}(\mathbf{r}) = -\frac{1}{4\pi} \sum_{i=1}^N \left( jk + \frac{1}{R_i} \right) \hat{\mathbf{R}}_i \times \mathbf{T}_i(\mathbf{r}) \quad (4)$$

$$\mathbf{T}_i(\mathbf{r}) = \int_{\Delta_i} \mathbf{J}_i(\mathbf{r}') \frac{e^{-jkR}}{R_i} ds' \quad (5)$$

Assume that the necessary transformations have been performed so that the incident field  $\mathbf{H}_s$  is given in terms of the reflector coordinate system. Then

$$\mathbf{J}_i(\mathbf{r}') = 2\hat{\mathbf{n}}_i \times \mathbf{H}_s(\mathbf{r}') \quad (6)$$

Next, assume that the incident field can be represented by a function of the form

$$\mathbf{H}_s = \mathbf{h}_s(\mathbf{r}_s) \frac{e^{-jk r_s}}{4\pi r_{si}} \quad (7)$$

where  $r_s$  is the distance to the source point. Then, Eq. (5) can be written

$$\mathbf{T}_i(\mathbf{r}) = \frac{\hat{\mathbf{n}}_i \times \mathbf{h}_s(\mathbf{r}_i)}{2\pi R_i r_{si}} \int_{\Delta_i} e^{-jk(R+r_s)} ds' \quad (8)$$

To simplify the form of the integration, the surface Jacobian is introduced within each triangular facet  $\Delta_i$ . For a planar surface  $z_i = f_i(x, y)$ , a normal is given by

$$\mathbf{N}_i = -\hat{x} f_{xi} - \hat{y} f_{yi} + \hat{z} \quad (9)$$

where

$$f_{xi} \equiv \frac{\partial f_i}{\partial x} \quad f_{yi} \equiv \frac{\partial f_i}{\partial y}$$

and a unit normal is given by

$$\hat{\mathbf{n}}_i = \frac{\mathbf{N}_i}{|\mathbf{N}_i|} \quad (10)$$

This permits the explicit evaluation of the Jacobian as

$$J_{\Delta_i} = |\mathbf{N}_i| = \left[ f_{xi}^2 + f_{yi}^2 + 1 \right]^{1/2} \quad (11)$$

Making use of the Jacobian then allows Eq. (8) to be rewritten as

$$\mathbf{T}_i(\mathbf{r}) = \frac{\hat{\mathbf{n}}_i \times \mathbf{h}_s(\mathbf{r}_i)}{2\pi R_i r_{si}} J_{\Delta_i} \int_{\Delta'_i} e^{-jk(R+r_s)} dx' dy' \quad (12)$$

in which  $\Delta'_i$  represents the area of the  $i$ th triangular facet projected onto the  $z = 0$  plane. Now, make a Taylor-series expansion of the exponent in Eq. (12). Retaining only the first-order terms, one can formally write

$$R(x, y) + r_s(x, y) = \frac{1}{k} (a_i - u_i x - v_i y) \quad (13)$$

in which  $a_i$ ,  $u_i$ , and  $v_i$  are constants. This approximation corresponds to a far-field approximation on the  $i$ th triangle. With this approximation, Eq. (12) reduces to

$$\mathbf{T}_i(\mathbf{r}) = \frac{\hat{\mathbf{n}}_i \times \mathbf{h}_s(\mathbf{r}_i)}{2\pi R_i r_{si}} J_{\Delta_i} e^{-j a_i} \int_{\Delta_i'} e^{j(u, x' + v, y')} dx' dy' \quad (14)$$

It may now be observed that this integral is the two-dimensional Fourier transform of the  $i$ th projected triangle  $\Delta_i'$ , expressed as

$$S(u, v) = \int_{\Delta_i'} e^{j(u, x + v, y)} dx dy \quad (15)$$

In order to explicitly evaluate the constants in Eq. (13), note that the equation of a plane can be expressed as

$$z = (x - x_i)f_{xi} + (y - y_i)f_{yi} + z_i$$

This can be used to obtain

$$a_i = kR(x_i, y_i) + kr_s(x_i, y_i) + u_i x_i + v_i y_i \quad (16)$$

$$\frac{u_i}{k} = \frac{(x_p - x_i) + (z_p - z_i)f_{xi}}{R(x_i, y_i)} + \frac{(x_s - x_i) + (z_s - z_i)f_{xi}}{r_s(x_i, y_i)} \quad (17)$$

$$\frac{v_i}{k} = \frac{(y_p - y_i) + (z_p - z_i)f_{yi}}{R(y_i, y_i)} + \frac{(y_s - y_i) + (z_s - z_i)f_{yi}}{r_s(y_i, y_i)} \quad (18)$$

Placing the result of Eq. (16) into Eq. (14), and recalling Eqs. (6) and (7), yields

$$\mathbf{T}_i(\mathbf{r}) = \mathbf{J}_i(r_i) J_{\Delta_i} e^{-j(u, x_i + v, y_i)} S(u_i, v_i) \frac{e^{-jkR_i}}{R_i} \quad (19)$$

This is the final form of the linear-phase approximation over each triangular facet. This expression can be used in Eq. (4) to compute the radiation integral once the Fourier transform of a triangular shape function  $S(u, v)$  is known. Fortunately, this transform can be computed in closed form [1] from the expression

$$S(u, v) = \sum_{n=1}^3 e^{j(u, x_n + v, y_n)} \frac{p_{n-1} - p_n}{(u + p_{n-1}v)(u + p_nv)} \quad (20)$$

in which  $(x_n, y_n)$  are the coordinates of the triangle vertices numbered in a clockwise direction. The slope of the  $n$ th side (between corners  $n$  and  $n+1$ ) is given by

$$p_n = \frac{y_{n+1} - y_n}{x_{n+1} - x_n} \quad (21)$$

Some attention must be given to the following special cases. First, if  $u = v = 0$ , the transform reduces to the formula for the area of a triangle

$$S(0, 0) = -\frac{1}{2} \left[ x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2) \right] \quad (22)$$

Next, if  $u/v \rightarrow -p_n$ , then

$$\begin{aligned} \lim_{u/v \rightarrow -p_n} S(u, v) &= \frac{p_{n+1} - p_{n-1}}{v^2(p_{n+1} - p_n)(p_{n-1} - p_n)} \\ &\times \left[ e^{j(u, x_{n-1} + v, y_{n-1})} - e^{j(u, x_n + v, y_n)} \right] \\ &+ \frac{(x_{n+1} - x_n)}{jv} e^{j(u, x_n + v, y_n)} \end{aligned} \quad (23)$$

### III. Numerical Results

A FORTRAN subroutine was written to perform the linear-phase calculations indicated above. Test cases were run for parabolas, ellipses, and shaped subreflectors, and the results were compared with the previous program, which used a constant magnitude and phase approximation on the triangular facets. A focused parabola is neither an interesting nor a challenging case for the algorithm, since the phase variation over the facet is small. As a simple test case, the far-field pattern and gain of a  $1,000\lambda$ -diameter parabolic reflector with a focal length of  $F = 400\lambda$  was calculated. The reflector is illuminated by a linearly polarized horn with a  $\cos\theta$  pattern function. Figure 2 compares the linear- and constant-phase approximation for a roughly equally spaced 80-by-80 rectangular grid of points divided into triangles over the reflector surface (approximately 10,000 triangles). The running time on a Cray X-MP was less than one minute. It has been previously demonstrated [3,4,5] that, once sufficient triangles to converge the solution have been utilized, the results of the constant-phase algorithm are valid, so only comparisons of the two techniques are presented.

A more interesting example is the ellipse shown in Fig. 3. The projected aperture of the ellipse is about 3 m, illumination function is a  $\cos^{42}\theta$  pattern function (22.3-dB gain), and the frequency is 31.4 GHz. The ellipse is about  $350\lambda$  along the major axis. Figure 4 compares the constant-phase approximation for different grid densities of approximately 4,000, 10,000, and 23,000 triangles and illustrates a general trend of the method, i.e., depending on the size of the triangles, there is an angular limit over

which the solution is valid. Figure 5 compares the linear-phase approximation with the constant-phase approximation for the 4,000-triangle case and demonstrates that the angular range is larger with the linear-phase approximation.

A third example is the shaped subreflector shown in Fig. 6. The diameter is 3.42 m (135 in.), and it is fed with a  $\cos^{233} \theta$  pattern function (29.7-dB gain). Figure 7 compares the results of a 4,000- and 10,000-triangle grid constant-phase approximation with a 4,000-triangle linear-phase approximation. The frequency of operation is 2.3 GHz, hence, the subreflector is about  $26\lambda$  in diameter. The 10,000-triangle constant phase is the converged result, and the 4,000-triangle linear case gives the same result. A very good approximation is also obtained with a 1,400-triangle grid for the linear case, but no meaningful results

are obtained with the constant-phase case. Figure 8 gives the linear-phase result for 31.4 GHz ( $360\lambda$  subreflector) using 23,000 triangles. No meaningful result is obtained for the equivalent constant-phase case.

Most of the examples given are for large reflectors to illustrate the robust character of the technique. For smaller reflectors ( $< 100\lambda$ ), meaningful results can be obtained on a PC in a reasonable time.

## IV. Conclusions

The linear-phase approximation is a significant improvement over the constant-phase approximation and enables the computation of fairly large (100 to  $1,000\lambda$ ) reflectors in a reasonable time (on the order of minutes) on a Cray computer.

## References

- [1] S. W. Lee and R. Mittra, "Fourier Transform of a Polygonal Shape Function and Its Application in Electromagnetics," *IEEE Trans. Antennas and Propagation*, vol. 31, no. 1, pp. 99–103, January 1983.
- [2] A. W. Rudge, K. Milne, A. D. Olver, and P. Knight, *The Handbook of Antenna Design, Vol. I*, London: Peter Peregrinus, 1982.
- [3] V. Galindo-Israel, T. Veruttipong, and W. Imbriale, "GTD, Physical Optics and Jacobi-Bessel Diffraction Analysis of Beamwaveguide Ellipsoids," *AP-S International Symposium*, Philadelphia, Pennsylvania, pp. 643–646, 1986.
- [4] T. Veruttipong, J. R. Withington, V. Galindo-Israel, W. A. Imbriale, and D. A. Bathker, "Design Considerations for Beamwaveguide in the NASA Deep Space Network," *IEEE Trans. Antennas and Propagation*, vol. AP-36, no. 12, pp. 1779–1787, December 1988.
- [5] A. G. Cha and W. A. Imbriale, "New Analysis of Beamwaveguide Antennas Considering the Presence of the Metallic Tube and Its Experimental Verification," *AP-S International Symposium*, Dallas, Texas, pp. 1506–1509, 1990.

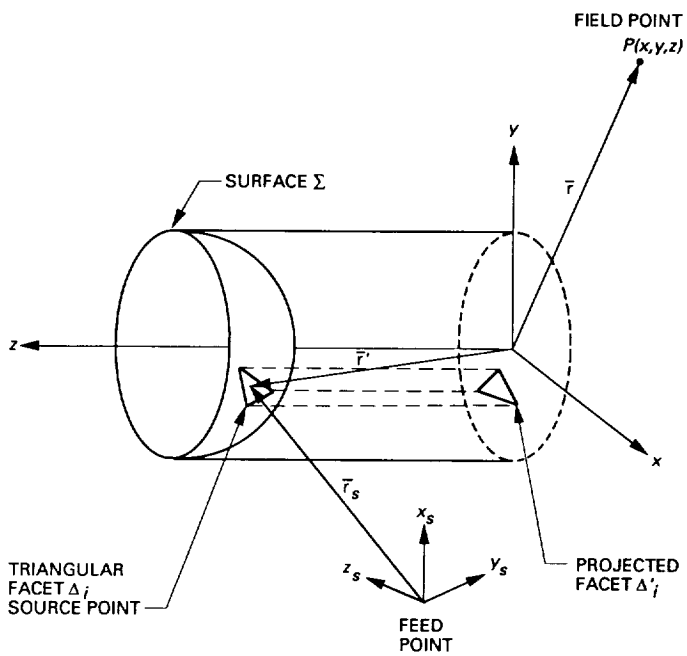


Fig. 1. Reflector analysis coordinate systems and a typical triangular facet.

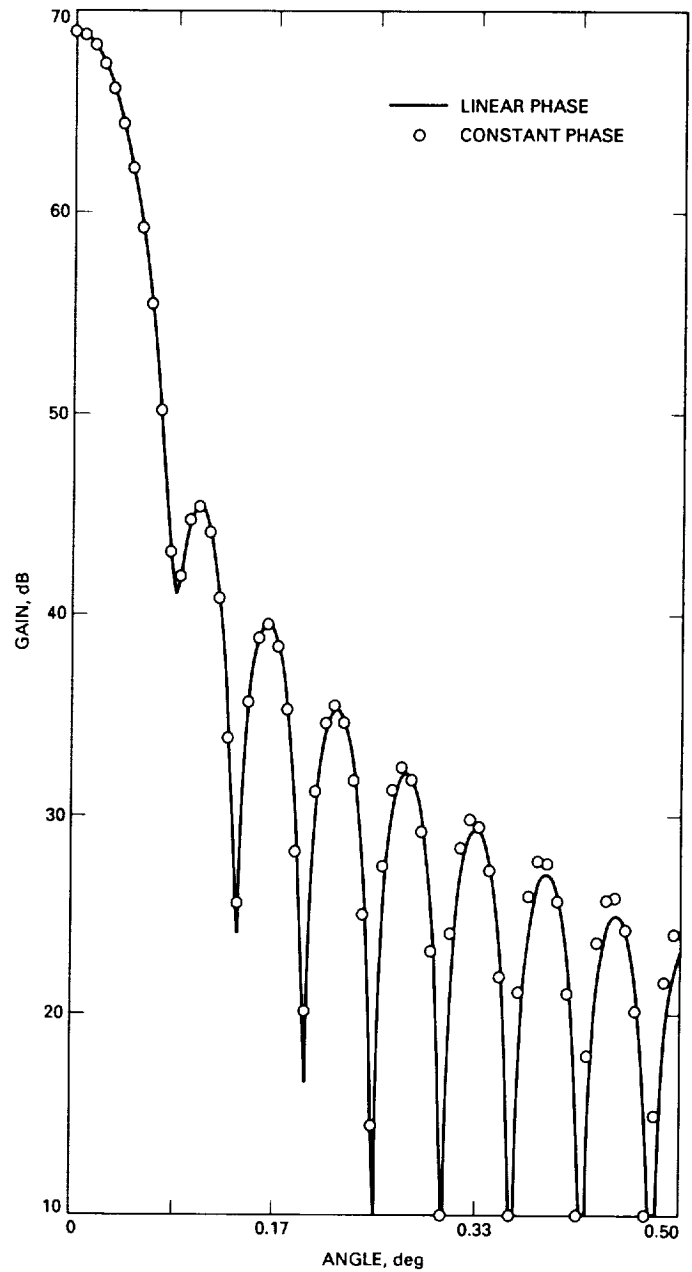
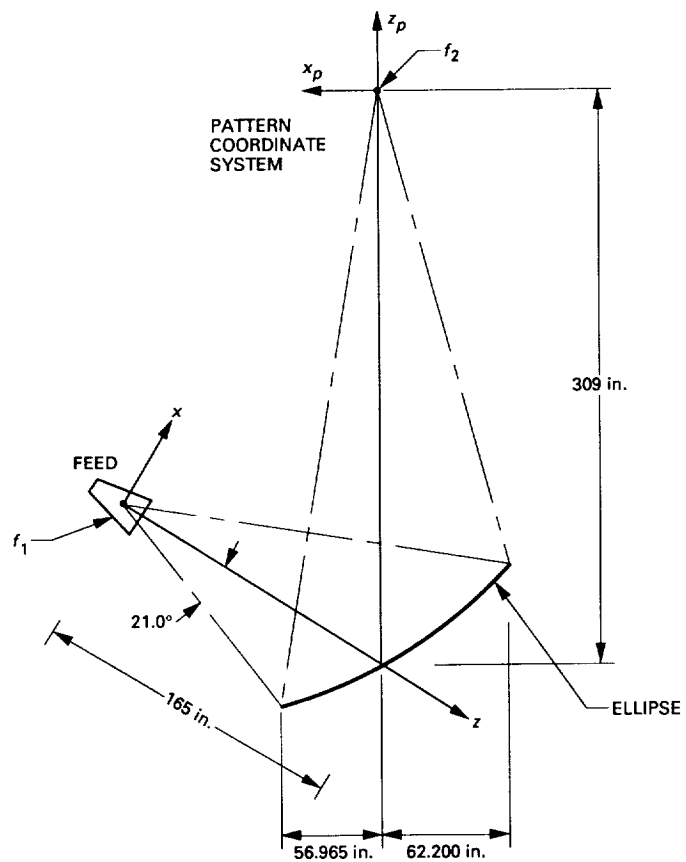


Fig. 2. Parabolic example.



**Fig. 3. Ellipse geometry.**

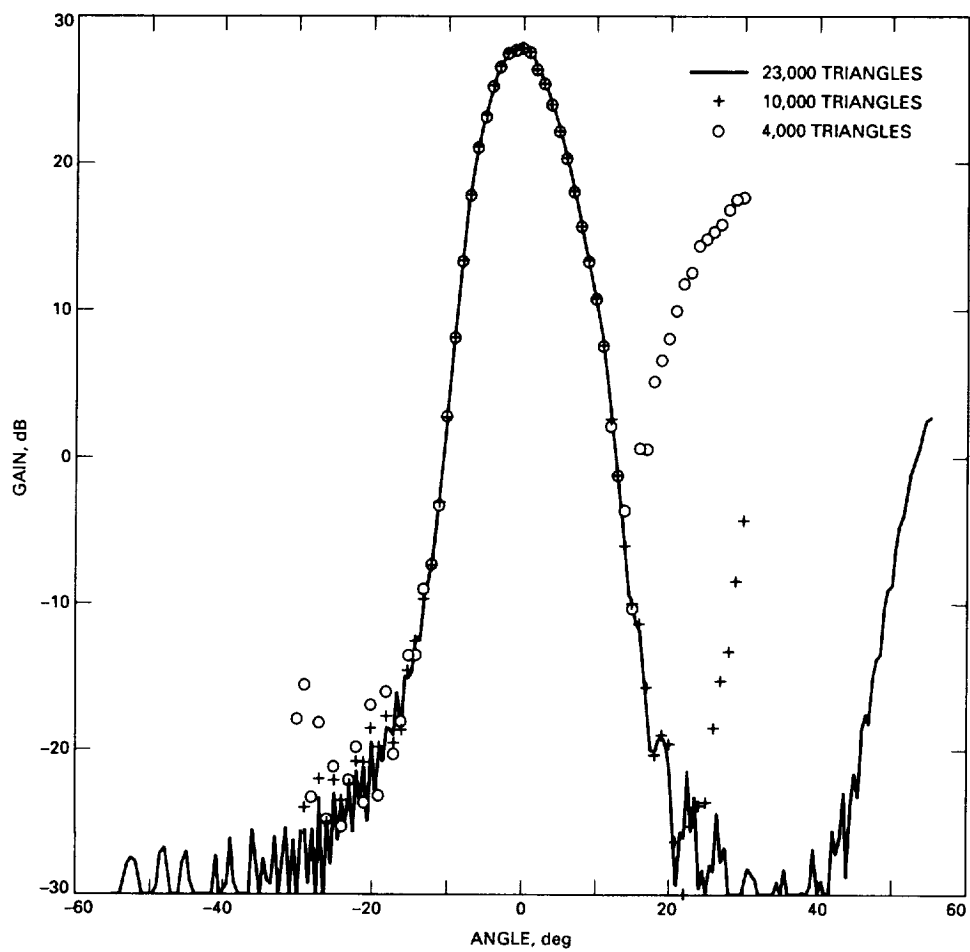


Fig. 4. Ellipse example: constant-phase approximation for offset plane.

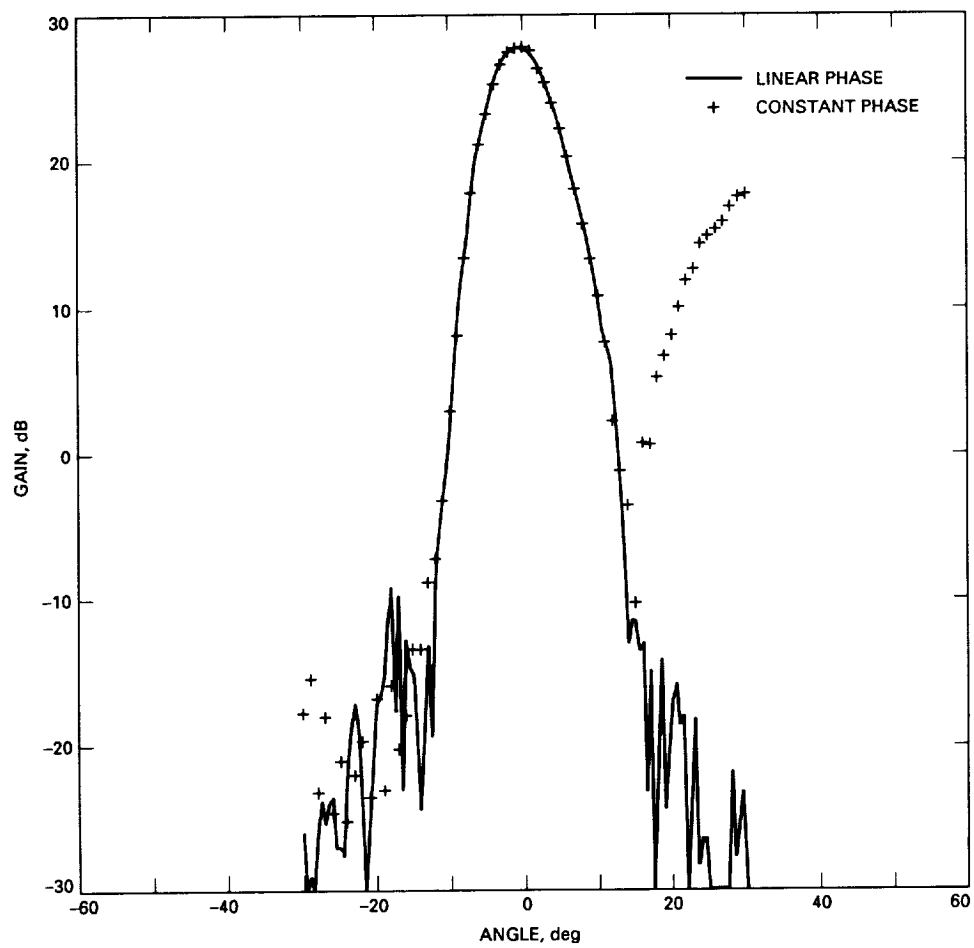


Fig. 5. Ellipse example: constant versus linear phase for offset plane.

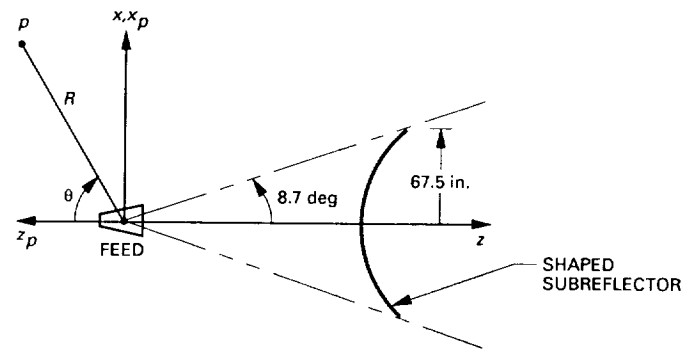


Fig. 6. Shaped subreflector geometry.

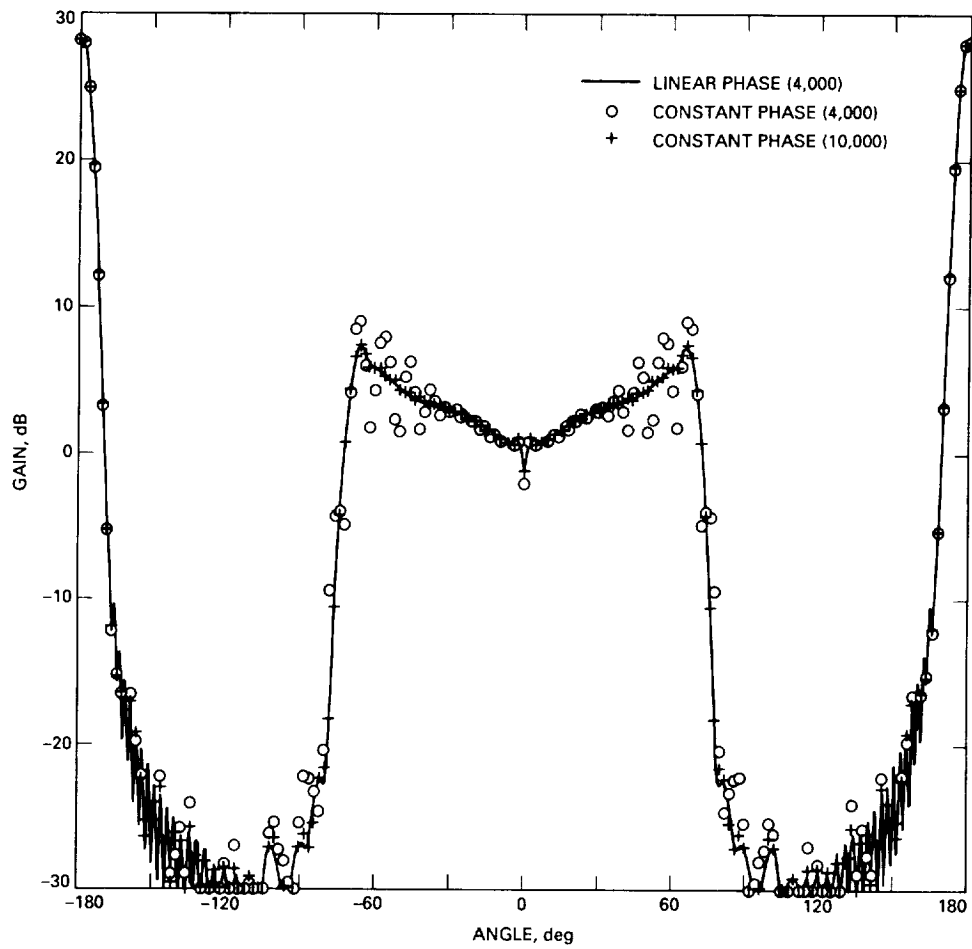


Fig. 7. Shaped subreflector example for H-plane at 2.3 GHz.

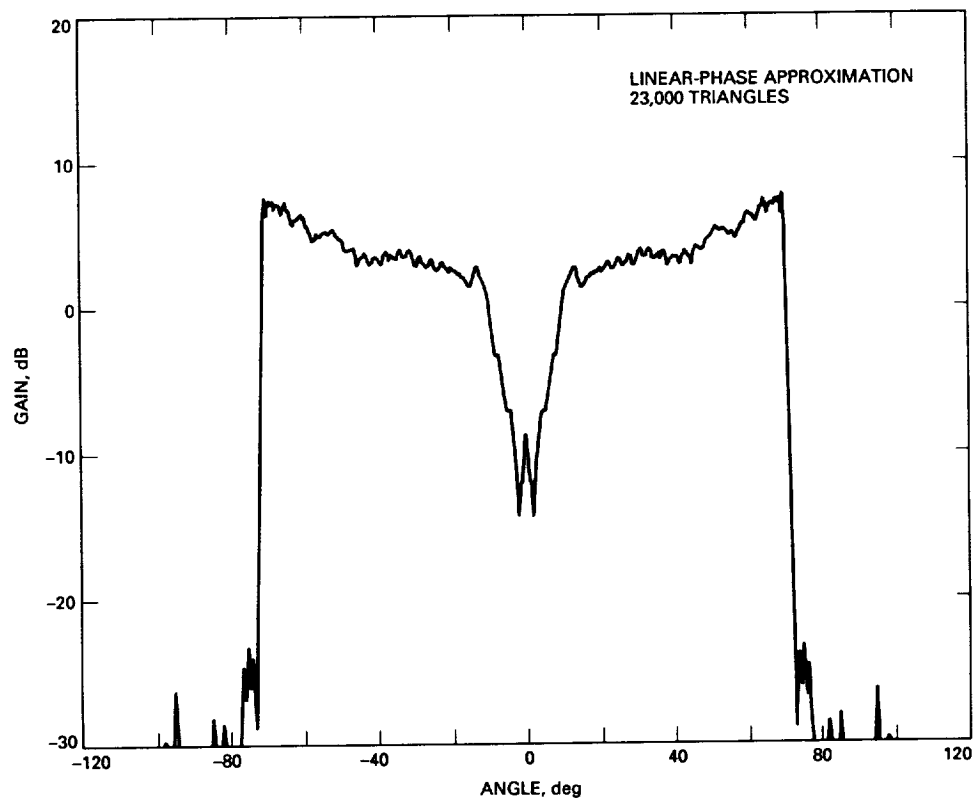


Fig. 8. Shaped subreflector, 31.4 GHz.

N91-11979

# Site Comparison for Optical Visibility Statistics in Southern Arizona

K. A. Cowles

Communications Systems Research Section

*One of the best locations in the continental United States for astronomical telescopes is southern Arizona. The mountains surrounding Tucson have clear skies 80 percent of the year, with image quality generally better than 2" on peaks. Two of the existing observatory sites in this area are being considered as locations for one of the three Atmospheric Visibility Monitoring (AVM) observatories. These sites are Mount Lemmon and Mount Hopkins. A comparison of the characteristics of each of the sites is made here to identify the more desirable of the two locations. It is recommended that Mount Lemmon be selected as the Arizona site for this project.*

## I. Introduction

The Atmospheric Visibility Monitoring (AVM) project is designed to create a model for cloud-cover correlation and atmospheric transmission in the southwestern United States for deep-space optical communications stations. By setting up three automatic telescopes to detect and measure light from selected stars, these atmospheric effects can be determined. A preliminary model has predicted a 94-percent probability of joint visibility from three sites located at least a few hundred kilometers apart in areas with at least 60 percent annual clear skies [1]. The mountains surrounding Tucson have clear skies 80 percent of the year, with image quality generally better than 2" on peaks [2]. This article describes the process for choosing a site in southern Arizona.

## II. Site Characteristics

Five major site characteristics were considered, each with a weighting for its importance to AVM. These char-

acteristics are the probability of clear skies (weighted at 30 percent), low particle scattering (20 percent), suitability for a future large optical-reception station (20 percent), availability of security and maintenance (15 percent), and low turbulence (15 percent) [3]. Each site was rated on a 1-to-10 basis for each characteristic. The ratings were then weighted and summed to give an overall rating for each site. This process is described below and was used to discriminate between Mount Lemmon and Mount Hopkins for the southern Arizona AVM site.

### A. Probability of Clear Skies

Mount Lemmon lies to the north of Tucson and Mount Hopkins to the south; they are within 130 km of each other. Satellite and ground statistics do not indicate a difference in the probability of clear skies for the two. Because Mount Lemmon is 0.5 km higher than Mount Hopkins, it is possible that Mount Lemmon may experience more clouds "clinging" to its peak at a higher altitude. However, it may also be above fog that can exist at the lower elevation of

Mount Hopkins. Local astronomers relate no appreciable difference in the number of cloudy or clear days between the two sites. Since they are both located in the area of the continental United States with the highest percentage of clear skies, both are rated as 9. A higher rating is not given because places such as Hawaii receive a higher annual percentage of clear skies.

## B. Low Particle Scattering

Low particle scattering is typically exhibited in areas with appreciable vegetation, which keeps dust down, and areas with little precipitable water vapor. Research at Mount Lemmon has shown precipitable water vapor to be very low (1.4 mm) [4]. Records of precipitable water vapor measurements at Mount Hopkins are not available. Vegetation is adequate at both locations for minimizing dust particles in the atmosphere. Winds from the southwest (the direction of Tucson) blow particles and smog toward Mount Lemmon. However, the peak of Mount Lemmon at 2.791 km [4] is protected by an inversion layer at 2 km, which confines valley haze and smog. This elevation also helps decrease particle scattering, since the atmosphere is less dense and there is less of it to look through. Mount Lemmon's higher elevation and reputation as a good infrared site make it rate slightly better (9) for scattering than Mount Hopkins (8).

## C. Suitability for a Future Large Optical Reception Station

Locating future 10-m receiving stations at AVM sites is also an important consideration in site selection. The AVM project will provide data on conditions that will be useful for such stations. Relations with the site management will be developed, and a working knowledge of the facilities and operations specific to a site will be gained. The management of Mount Lemmon is very receptive to the addition of a larger optical station, and there are several large areas where one could be located. There is a plateau at the peak, whereas Mount Hopkins has a ridge; Mount Lemmon thus has more space available. Mount Hopkins has a limited amount of space for large structures and, at this time, the management of Mount Hopkins would not approve the addition of a large telescope for communications. It is reserving the remaining sites for projects in astronomy. Five years from now the management might change its position, but there are no predictions.<sup>1</sup> Mount Hopkins is therefore rated very low for future suitability (2); Mount Lemmon is rated high (10).

<sup>1</sup> From a conversation with John Huchra, Center for Astrophysics, Cambridge, Massachusetts, November 14, 1989.

## D. Availability of Security and Maintenance

The availability of security and maintenance at a site is also a consideration in order to make frequent visits by JPL personnel feasible. Security is necessary to protect the observatories from vandalism and prevent accidents caused by hikers who might get in the way of a moving telescope or roof. Both Mount Hopkins and Mount Lemmon are surrounded by gates and do not experience a large number of visitors. An additional fence around the AVM enclosure may be necessary at any site.

Necessary minimal maintenance would include snow removal and checks for weather damage and mechanical failure. It would also be useful to have someone at the site to receive calls from JPL personnel and respond appropriately to problems detected. The frequency of required periodic checks has not been determined.

Mount Lemmon has personnel available for daily checks on all the buildings on the peak. While these people are willing to make minor adjustments, they prefer not to be responsible for solving large problems. Mount Hopkins, however, has limited human resources. Support personnel are on the mountain from 9 a.m. to 4 p.m. weekdays. Two people are usually there, although additional people are available if snow needs to be plowed. However, the management of Mount Hopkins is hesitant to add to the duties of the support people.<sup>2</sup> On clear nights, only observers are on the mountain, and no one is on the mountain on cloudy nights. For the months of July and August, the observatories are closed for the rainy season.

Lou Boyd of Fairborn Observatory is in charge of maintaining the automatic telescopes for the Automatic Photoelectric Telescope Service on Mount Hopkins. He has offered to make routine checks of those AVM observatories located near Fairborn on Mount Hopkins. He is on the mountain only two or three days a week, but he is very familiar with automatic telescopes. He could fix any problem with the hardware, but he prefers not to have to change any software. This should not be a problem.

Mr. Boyd gets \$15,000 a year from Vanderbilt University for his services, and would consider a similar arrangement with JPL. He would not be available in the case of an emergency; in fact, no one at Mount Hopkins would be available for such a response.

Since both sites have about the same degree of security, and maintenance would be available at all times on Mount

<sup>2</sup> *ibid.*

Lemmon, Mount Hopkins was rated lower (5) than Mount Lemmon (7).

### E. Turbulence

Turbulence causes blurring of an image. Sites with low turbulence usually have a steep slope that rises from a valley in the direction of the wind source. Also, for low turbulence, there should be no peaks in that direction higher than 1 km for at least 30 km.

In Tucson, the winds come from the west and southwest for most of the year. Mount Lemmon is the highest peak in the Santa Catalina Mountains and has a medium slope up from lower parts of the range; the top of Mount Lemmon is a plateau. Mount Hopkins has a ridge geometry, which produces lower turbulence.

Turbulence is often measured by rms image motion. An average measurement of about 2" is considered excellent for optical communications. The rms image motion averages 2.1" on Mount Lemmon [5]. Typical rms image motion measurements are 0.6" at Mount Hopkins [6]. (Mauna Kea exhibits rms image motion  $\leq 1''$  only 33 percent of the nights [7].) Mount Hopkins is therefore rated higher (10) for low turbulence than Mount Lemmon (8).

### III. Rating

The ratings for Mount Lemmon and Mount Hopkins in each characteristic are given in Table 1.

There is also a cost consideration. The Steward Observatory, University of Arizona, which runs Mount Lemmon, has given a cost estimate for services as shown in Table 2.

The total fee includes daily site security checks, winter snow removal, general (light) maintenance, and limited use of the machine shop and dormitory facilities. Not included are electricity, telephone, or installation charges. The Steward Observatory explained that this is a fair share of the costs necessary to keep the mountain operating and includes road and power-line maintenance.

Mount Hopkins would not charge for site rental, but would not provide much in the way of service. Maintenance at Mount Hopkins performed by Lou Boyd would cost \$15,000 annually.

### IV. Conclusions

Both Mount Lemmon and Mount Hopkins are high-quality astronomical sites. Based on this comparison study, it was recommended that Mount Lemmon be selected for the Arizona AVM site. The cost of operating at Mount Lemmon is a drawback, but it will cost less than \$6,000 a year more than at Mount Hopkins, and the benefits outweigh this cost. The minimum amounts of maintenance provided will be adequate since the observatories are autonomous. Contract negotiations with Mount Lemmon have begun.

## References

- [1] K. Shaik, "A Preliminary Weather Model for Optical Communications Through the Atmosphere," *TDA Progress Report 42-95*, vol. July-September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 212-218, November 15, 1989.
- [2] *Climatic Atlas of the United States*, U.S. Department of Commerce, Washington, D.C., 1968.
- [3] K. Cowles, "Site Selection Criteria for the Optical Atmospheric Visibility Monitoring Telescopes," *TDA Progress Report 42-97*, vol. January-March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 235-239, May 15, 1989.
- [4] G. P. Kuiper, "High Altitude Sites and IR Astronomy, II," *Lunar and Planetary Laboratory Communications* No. 156, U. of Arizona, Tucson, pp. 337-382, 1970.
- [5] G. V. Coyne, "Preliminary Report on Optical Seeing Tests at Mount Lemmon, March-June 1971," *Lunar and Planetary Laboratory Communications* No. 170, U. of Arizona, Tucson, pp. 185-189, 1971.
- [6] M. R. Pearlman, et al., "SAO/NASA Joint Investigation of Astronomical Viewing Quality at Mount Hopkins Observatory: 1969-1971," *SAO Special Report 357*, Cambridge, Massachusetts: Smithsonian Astrophysical Observatory, January 1, 1974.
- [7] H. M. Dyck, and R. R. Howell, "Seeing Measurements at Mauna Kea from Infrared Speckle Interferometry," *Pub. Astr. Soc. Pac.*, vol. 95, pp. 786-791, October 1983.

**Table 1. Ratings**

Characteristic, weight	Mount Lemmon	Mount Hopkins
Clear skies (30 percent)	9	9
Low scattering (20 percent)	9	8
Future suitability (20 percent)	10	2
Security and maintenance (15 percent)	7	5
Turbulence (15 percent)	8	10
Rating	8.75	6.95

**Table 2. Annual expenditures at Mount Lemmon**

Item	Fee
10 percent resident engineer salary	\$4,039
10 percent forest service lease	\$6,500
10 percent site maintenance costs	\$9,800
Total	\$20,339

732  
305345

1-11

# The Theoretical Limits of Source and Channel Coding

S. J. Dolinar and F. Pollara  
Communications Systems Research Section

*This article presents the theoretical relationship among signal power, distortion, and bandwidth for several source and channel models. It is intended as a reference for the evaluation of the performance of specific data compression algorithms.*

## I. Introduction

The theoretical limits on the performance of source and channel coding are well known for several source and channel models [1,3]. In this article these limits are calculated for the Gaussian channel, used as a model of the deep-space channel, and for some simple sources, potentially useful as models of planetary images. The formulas underlying these calculations are well known; the aim of this article is to collect and graphically display the results. The performance of specific data compression algorithms can be compared to the ultimate limits shown in these graphs. Similar results were presented in [2] for a binary symmetric source with probability of error distortion criterion.

The results show the basic tradeoffs that must be made among signal power, distortion, and bandwidth, assuming that the system designer is free to design encoders and decoders but has no control over the source, the channel, or the user. These results also suggest that improvements in information transmission in future missions should be sought primarily through better source encoding rather than by pushing channel coding gain closer to its limit.

## II. The Communication System

Consider a communication system as shown in Fig. 1. The source produces *source symbols*  $x$  with average en-

ergy  $\mathcal{E}_x$  at the rate of  $R_x$  source symbols per second. The source encoder outputs bits  $b$  with average energy  $\mathcal{E}_b$  at the rate of  $R_b$  bits/sec. The channel encoder produces *channel symbols*  $y$  with average energy  $\mathcal{E}_y$  at the rate of  $R_y$  channel symbols per second. The channel produces noisy outputs  $\hat{y}$ ; the channel decoder and source decoder produce outputs  $\hat{b}$  and  $\hat{x}$ , attempting to reproduce the values of  $b$  and  $x$ , respectively.

### A. Channel Models

Channel capacity is defined as

$$C(\cdot) = \max I(y; \hat{y}) \quad (1)$$

where  $I(y; \hat{y})$  is the average mutual information provided by the output  $\hat{y}$  about the input  $y$ . The maximization is performed over all probability densities  $q(y)$  satisfying the desired cost constraint  $E[y^2] \leq \mathcal{E}_y$ . The capacity is written as a function of an unspecified argument; it depends on the energy constraint  $\mathcal{E}_y$  and the transition statistics  $p(\hat{y}|y)$  that characterize the channel.

**Case C1.** For a discrete-time additive white Gaussian noise channel with continuous amplitude input and output, and with noise distribution  $N(0, \mathcal{E}_n)$ , Eq. (1) yields the capacity function in bits per channel symbol [3]:

$$C(\rho) = \frac{1}{2} \log_2(1 + \rho)$$

where  $\rho = \mathcal{E}_y/\mathcal{E}_n = 2\mathcal{E}_y/N_0$ , and  $N_0/2$  is the two-sided spectral density of the noise.

As a comparison consider two additional channel models obtained by constraining the input and/or the output of the additive white Gaussian noise channel. One channel, Case C3, is obtained by restricting the channel input to binary symbols, while leaving the channel output unconstrained. The other case, Case C2, is obtained by constraining the channel output as well as its input to be binary, that is, by hard-quantizing the Gaussian channel output to two levels.

**Case C2.** A hard-quantized Gaussian channel with binary signaling and signal levels  $\pm\sqrt{\mathcal{E}_y}$  is a binary symmetric channel (BSC) with crossover probability  $\epsilon = Q(\sqrt{\rho})$ , where  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du$ . The capacity for such a channel in bits per channel symbol is [3]

$$C(\rho) = 1 - \mathcal{H}[Q(\sqrt{\rho})]$$

where  $\mathcal{H}(\epsilon) = -\epsilon \log_2 \epsilon - (1 - \epsilon) \log_2 (1 - \epsilon)$  is the binary entropy function.

**Case C3.** For the binary input Gaussian channel, the capacity can be written [2]

$$C(\rho) = 1 - E_u[\log_2(1 + e^{-2u})]$$

where  $u$  is a random variable with distribution  $N(\rho, \rho)$ , and  $E_u$  represents expectation over  $u$ .

## B. Source Models

A discrete-time, continuous-amplitude stationary source is considered. Its rate distortion function is defined as

$$R(\cdot) = \min I(x; \hat{x})$$

The minimization is over all conditional probability densities  $p(\hat{x}|x)$  satisfying a distortion constraint  $E[d(x, \hat{x})] \leq D$ , where  $d(\cdot, \cdot)$  is a distortion measure. As with channel capacity,  $R(\cdot)$  is written as a function of an unspecified argument; it depends on the distortion constraint  $D$  and on the source statistics  $p(x)$ .

**Case S1.** For a Gaussian memoryless source with mean square error (MSE) distortion constraint  $E[z^2] \leq \mathcal{E}_z$ , where  $z = \hat{x} - x$ , the rate distortion function can be expressed in bits per source symbol as [1]

$$R(\delta) = \frac{1}{2} \log_2(1/\delta), \quad 0 < \delta < 1 \quad (2)$$

where  $\delta = \mathcal{E}_z/\mathcal{E}_x$  is the normalized MSE distortion.

**Case S2.** As a comparison consider a memoryless binary symmetric source (BSS) under the probability of error distortion measure,  $\text{Prob}(\hat{x} \neq x) \leq P_e$ . This source has the rate distortion function [1]

$$R(P_e) = 1 - \mathcal{H}(P_e), \quad 0 \leq P_e \leq 1/2$$

**Case S3.** A simple example of a source with memory is a stationary Gaussian source with MSE distortion constraint as in Case S1. In this case only a parametric form for  $R(\delta)$  is known [1]:

$$\delta(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min[\theta, \Phi_0(\omega)] d\omega$$

$$R(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max[0, \log_2 \frac{\Phi_0(\omega)}{\theta}] d\omega$$

where  $\Phi_0(\omega) = \sum_{k=-\infty}^{\infty} \frac{\phi(k)}{\phi(0)} e^{-jk\omega}$ , and  $\frac{\phi(k)}{\phi(0)}$  is the normalized autocorrelation function of the Gaussian process. Of particular interest is the case of a first-order Markov source with autocorrelation  $\phi(k) = \mathcal{E}_x \gamma^{|k|}$ ,  $0 < \gamma < 1$ , which provides a good approximation to the autocorrelation function of actual images, as shown in the example of Fig. 2. The rate distortion function for a Gauss-Markov source is shown in Fig. 3 for several values of  $\gamma$ . Note that, at any given distortion level, the rate  $R$  varies insignificantly with  $\gamma$  for moderate values of the correlation coefficient ( $\gamma < 0.5$ ) but is reduced dramatically as  $\gamma$  approaches 1.

## III. The Source-Channel Coding Theorem

Shannon's channel coding theorem and source coding theorem can be merged into the source-channel coding theorem [3], which is the central result of information theory. The source-channel coding theorem answers the following fundamental question: Given a source with rate distortion function  $R(\cdot)$  and a channel with capacity function  $C(\cdot)$ , under what conditions is it possible, with sufficient coding, to achieve source distortion and channel cost that do not exceed the constraints used in the definitions of  $R(\cdot)$  and  $C(\cdot)$ , respectively? According to the theorem, this is always possible if the capacity  $C(\cdot)$  is greater than the rate distortion  $R(\cdot)$  (where the two functions must be measured in consistent units), and never possible if the rate distortion is greater than the capacity.

In general, the source symbol rate  $R_x$  need not correspond to the channel symbol rate  $R_y$ . Therefore the requirement on the information transmission system is normalized as

$$C(\cdot)R_y > R(\cdot)R_x$$

or

$$C(\cdot) > rR(\cdot) \quad (3)$$

where  $r = R_x/R_y$  is the number of source symbols per channel symbol. A large  $r$  corresponds to a small channel bandwidth requirement relative to the bandwidth of the source. To get an idea of the practical range of interest of  $r$ , consider the following examples: A system with an 8-bit/symbol source and no source compression on a channel encoded with code rate 1/6 yields  $r \simeq 0.02$ ; at another extreme, the same source with a source compression ratio of 8:1 and no channel coding yields  $r = 1$ ; the Voyager concatenated system with a source compression of 2:1 yields  $r \simeq 0.11$ .

This requirement can now be applied to all the combinations of source and channel models introduced in Section II. For all three channels considered in Section II, the capacity  $C(\cdot)$  is a function of a single variable  $\rho$ , which is a signal-to-noise ratio involving the energy per encoded channel symbol. Usually it is more meaningful to measure the channel cost in terms of a signal-to-noise ratio,  $\rho_b = \mathcal{E}_b/\mathcal{E}_n = 2\mathcal{E}_b/N_0$ , involving the energy per channel encoder input bit, or a signal-to-noise ratio,  $\rho_x = \mathcal{E}_x/\mathcal{E}_n = 2\mathcal{E}_x/N_0$ , involving the energy per source symbol. The variables  $\rho$ ,  $\rho_b$ , and  $\rho_x$  are related by

$$r\rho_x = \frac{R_x}{R_y}\rho_x = \rho = \frac{R_b}{R_y}\rho_b = r_c\rho_b = rr_s\rho_b$$

where  $r_c = R_b/R_y$  is the rate of the channel code, and  $r_s = R_b/R_x$  is the rate of the source code. Therefore the capacity constraint, Eq. (3), can be equivalently written in terms of  $\rho_x$  as

$$C(r\rho_x) > rR(\cdot)$$

or, in terms of  $\rho_b$  as

$$C(rr_s\rho_b) > rR(\cdot)$$

With  $r_s$  taken to equal its minimum allowable value,  $R(\cdot)$ , the latter requirement is

$$C(rR(\cdot)\rho_b) > rR(\cdot)$$

These formulas give the required relationships among  $r$ ,  $\rho_b$  or  $\rho_x$ , and the unspecified distortion variable in  $R(\cdot)$ .

**Case S1.C1.** A Gaussian channel and a Gaussian memoryless source under the MSE distortion measure gives the requirement

$$\rho > \delta^{-r} - 1 \quad (4)$$

For a given source and channel the designer seeks a realization with small  $\delta$  and  $\rho$ , and large  $r$ . Of course these are conflicting requirements, limited by Eq. (4), and illustrated in Figs. 4(a), 4(b), and 4(c), where the feasible region is above each curve with a given parameter  $r$ . When the source and channel have the same clock rate ( $r = 1$ ) and  $\rho + 1 = 1/\delta$ , the source and channel are perfectly matched, and ideal performance  $C = R$  is obtained without any need for source or channel coding, assuming proper scaling of the signal levels. If the channel clock rate is much higher than the source clock rate, the required  $\mathcal{E}_b/N_0$  approaches the well known infinite bandwidth expansion limit of  $-1.59$  dB.

**Case S2.C1.** For a BSS and an unconstrained Gaussian channel the following requirement, shown in Figs. 5(a), 5(b), and 5(c), is obtained:

$$\rho > 2^{2r[1-\mathcal{H}(P_e)]} - 1$$

In this case the distortion  $P_e$  is extremely sensitive to small changes in  $\mathcal{E}_b/N_0$  or  $r$ . Conversely, extremely low error probabilities can be obtained as long as  $\mathcal{E}_b/N_0$  is greater than the channel capacity limit for any fixed channel code rate,  $r_c = r$ .

**Case S1.C2.** A Gaussian memoryless source and a BSC gives the requirement

$$2^{[1-\mathcal{H}(\epsilon)]} > \delta^{-r/2}$$

which relates the channel cost  $\epsilon = Q(\sqrt{\rho})$  to the normalized MSE distortion  $\delta$ , as in Figs. 6(a), 6(b), and 6(c). The asymptotic behavior of these graphs for large signal-to-noise ratio is due to the hard quantization introduced by the BSC. The quantization loss can be seen by comparing Fig. 6(a) to Fig. 4(a), Fig. 6(b) to Fig. 4(b), or Fig. 6(c) to Fig. 4(c). The required  $\mathcal{E}_b/N_0$  at infinite bandwidth expansion is 0.37 dB, or 1.96 dB higher than the corresponding requirement for the unquantized channel.

**Case S2.C2.** For a BSS and BSC under the  $P_e$  distortion measure, the requirement is

$$1 - \mathcal{H}(\epsilon) > r[1 - \mathcal{H}(P_e)]$$

as illustrated in Figs. 7(a), 7(b), and 7(c). When  $r = 1$  and  $\epsilon = P_e$ , another case of perfect matching between source and channel is obtained, which yields ideal performance  $C = R$  without any need for source or channel coding. The loss due to the hard quantization of the channel can be seen by comparing Fig. 7(a) to Fig. 5(a), or Fig. 7(b) to Fig. 5(b), or Fig. 7(c) to Fig. 5(c).

**Case S2.C3.** The cases involving the binary input Gaussian channel do not yield a simple analytical expression. The results are shown in Figs. 8(a), 8(b), and 8(c), for this channel combined with the BSS. This source-channel combination is the same case considered in [2], and Fig. 8(c) is directly comparable to Fig. 1 of [2]. (Note that  $E_b/N_0$  in [2] refers to the source symbol signal-to-noise ratio denoted  $\mathcal{E}_x/N_0$  in this article.)

**Case S3.C1.** Finally, the case of a Gauss-Markov source and Gaussian channel is shown in Figs. 9, 10, and 11, for three selected values of the rate  $r$  and several values of the correlation coefficient  $\gamma$ . It is apparent that a Gauss-Markov source with  $\gamma = 0.99$ , which is a good model for some highly correlated planetary images, offers large savings in signal power due to the high redundancy of the source.

## IV. Conclusion

The results presented provide a reference against which the performance of specific data compression algorithms can be measured. Choosing an appropriate model from Figs. 4–11, and a desired ratio  $r$  of source symbol rate to channel symbol rate, one can determine the minimum channel symbol signal-to-noise ratio  $\mathcal{E}_y/N_0$ , bit signal-to-noise ratio  $\mathcal{E}_b/N_0$ , or source symbol signal-to-noise ratio  $\mathcal{E}_x/N_0$  required to produce a normalized source distortion that is less than  $\delta$  or  $P_e$ .

## References

- [1] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, New Jersey: Prentice Hall, 1971.
- [2] S. A. Butman and R. J. McEliece, "The Ultimate Limits of Binary Coding for a Wideband Gaussian Channel," *DSN Progress Report 42-22*, vol. May–June 1974, Jet Propulsion Laboratory, Pasadena, California, pp. 78–80, August 15, 1974.
- [3] R. J. McEliece, *The Theory of Information and Coding*, Reading, Massachusetts: Addison-Wesley, 1977.

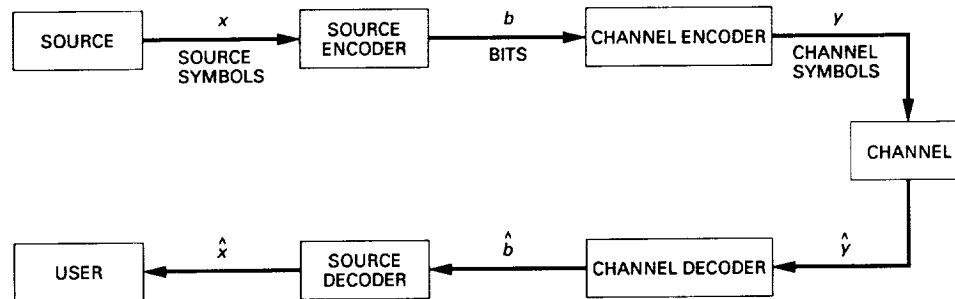


Fig. 1. A communication system model.

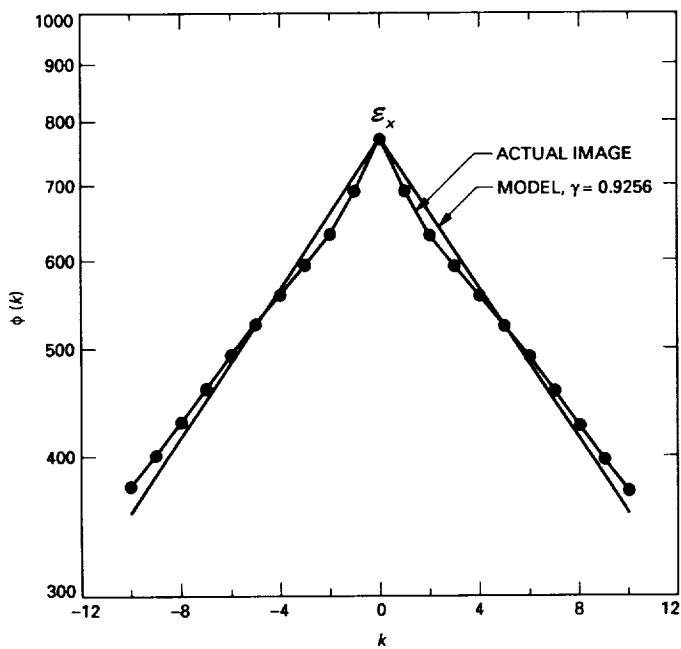


Fig. 2. Autocorrelation functions of a first-order Markov model and of an image of the moon.

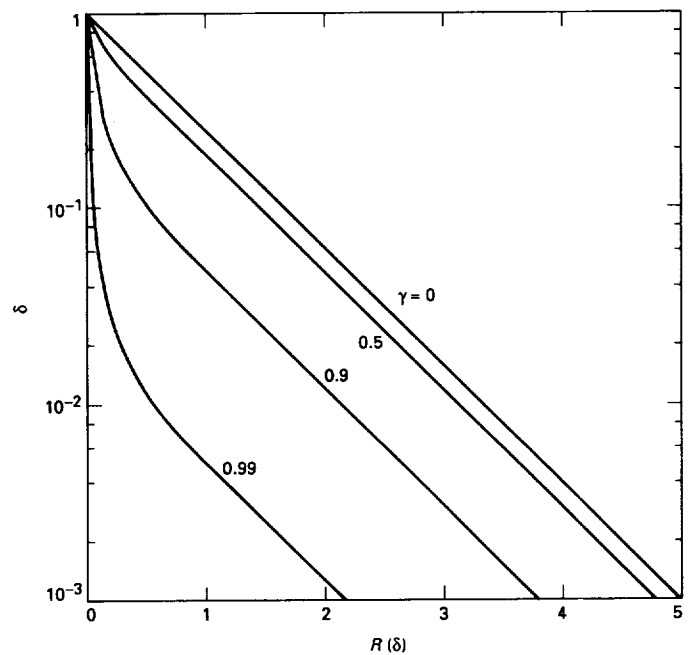


Fig. 3. Rate distortion function for a Gauss-Markov source with correlation  $\gamma$ .

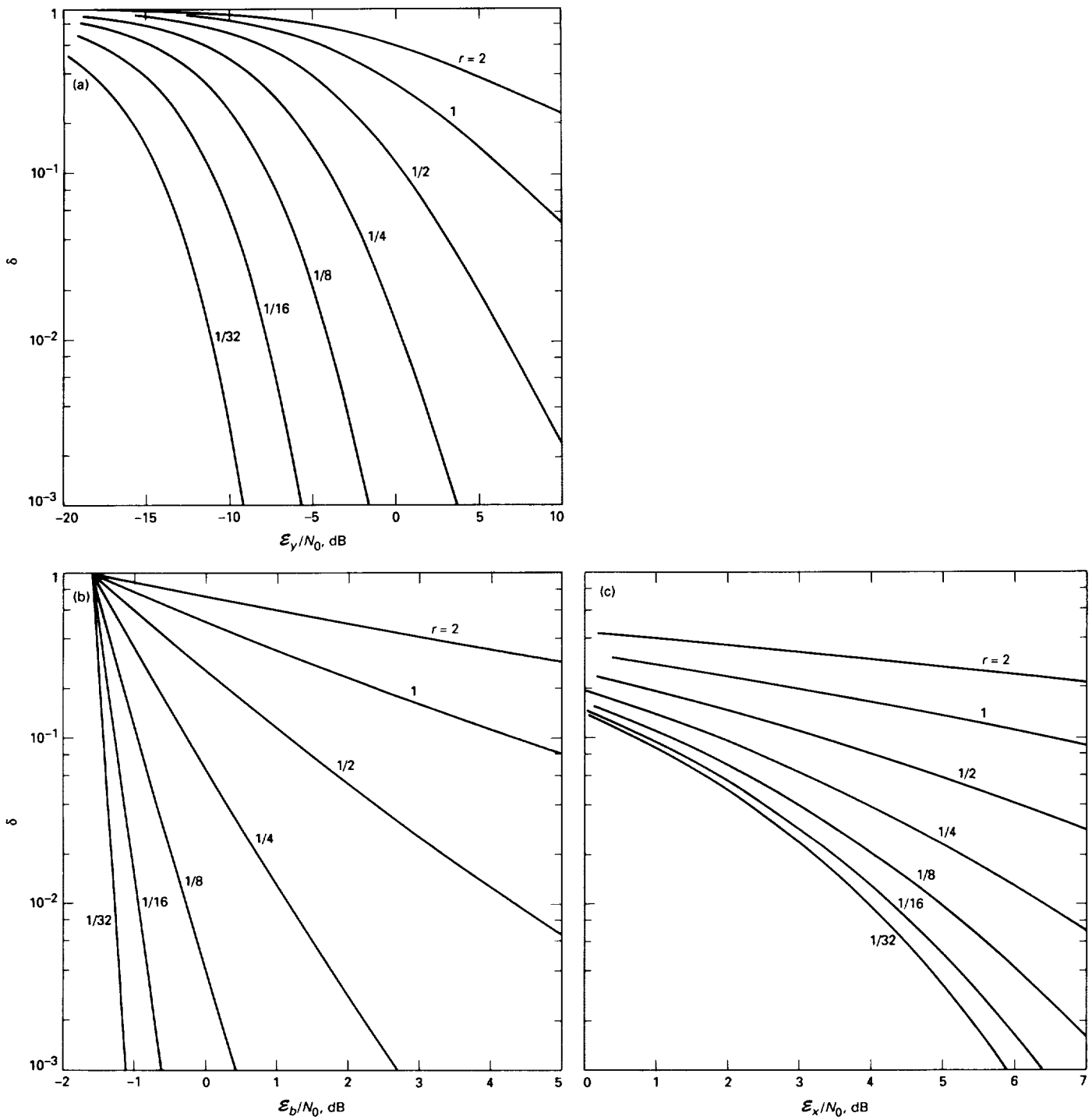


Fig. 4. Source-channel coding limits for a Gaussian source and a Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

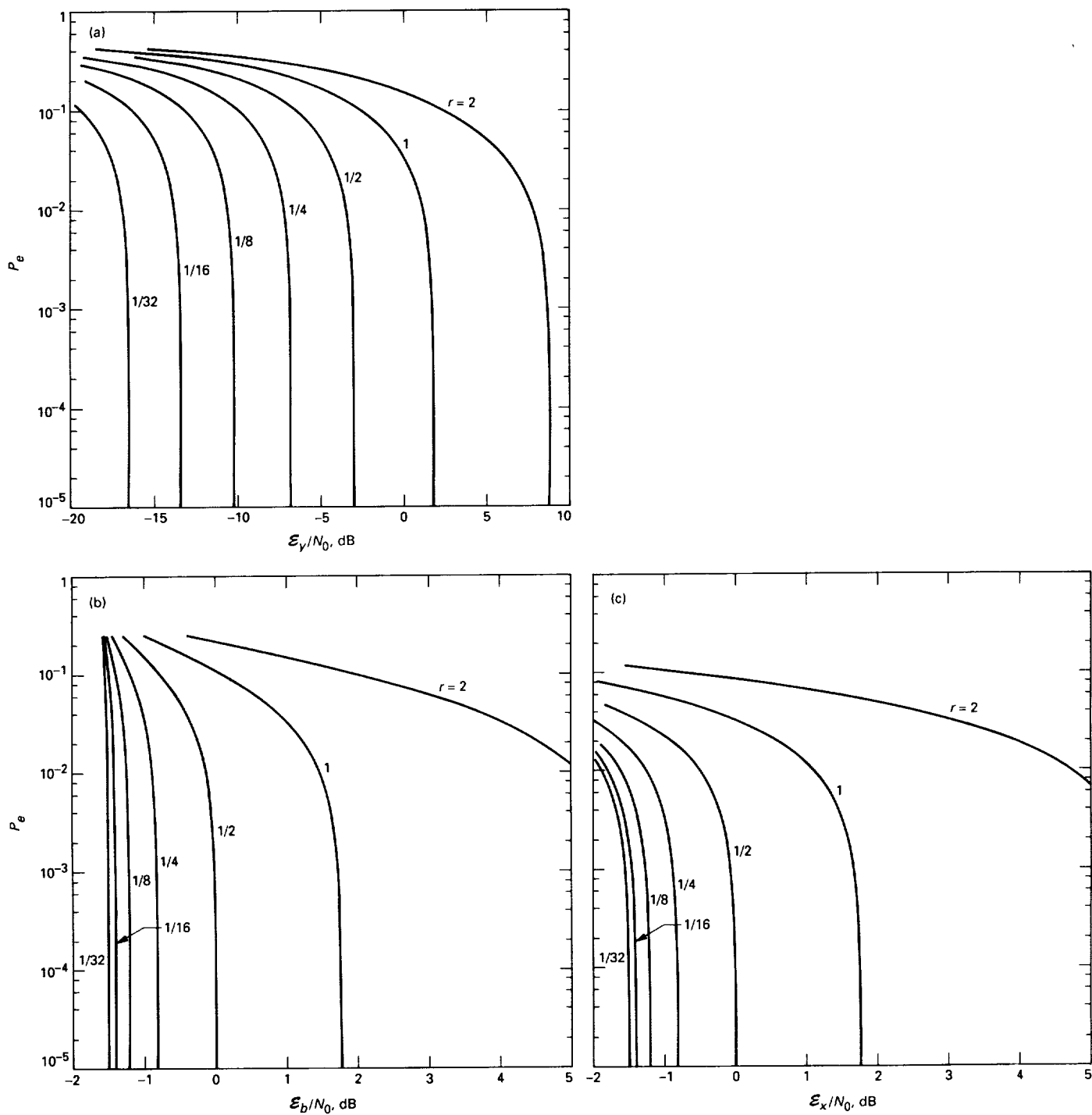


Fig. 5. Source-channel coding limits for a binary symmetric source (BSS) and a Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

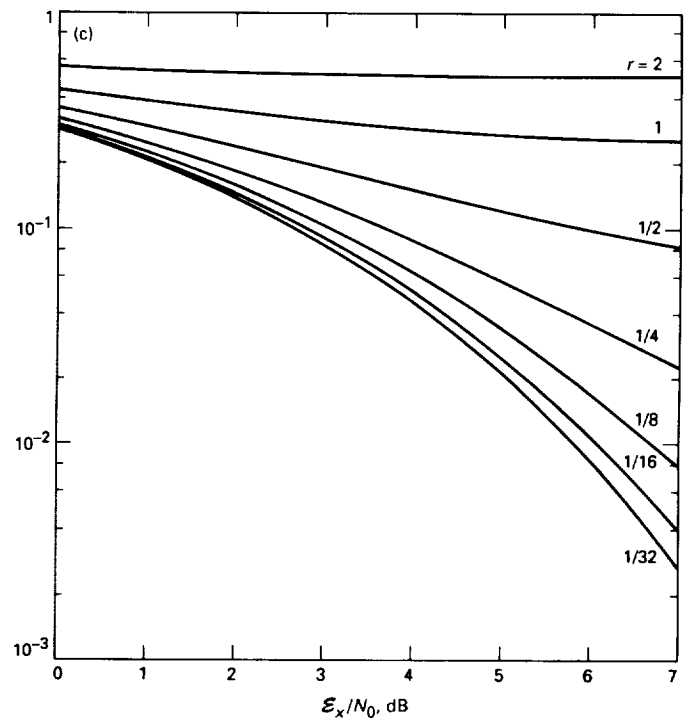
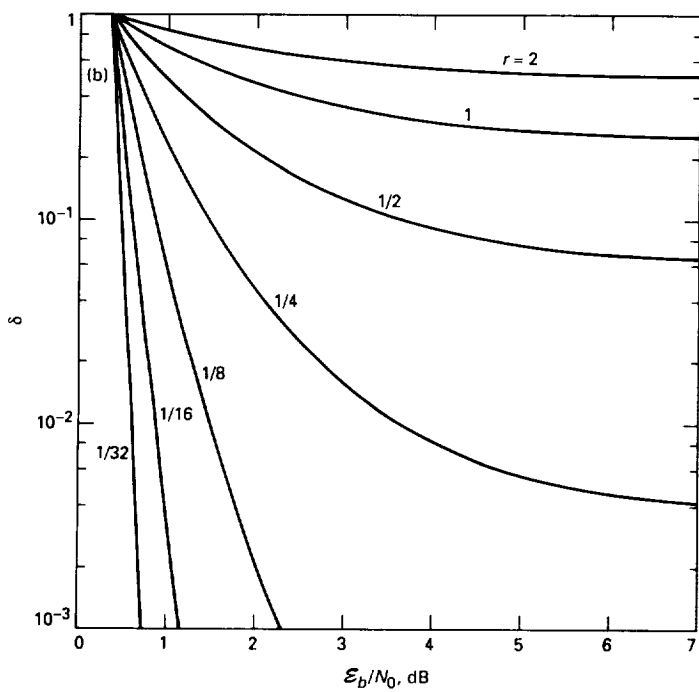
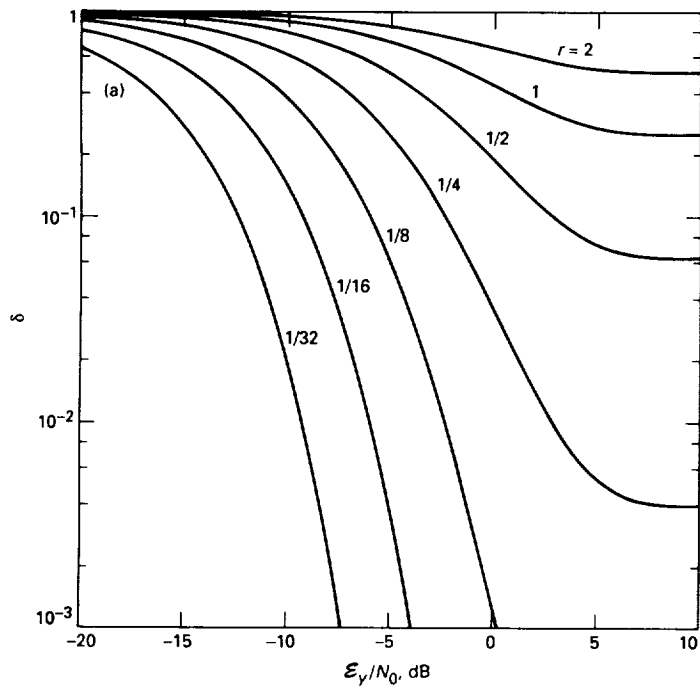


Fig. 6. Source-channel coding limits for a Gaussian source and a hard-quantized Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

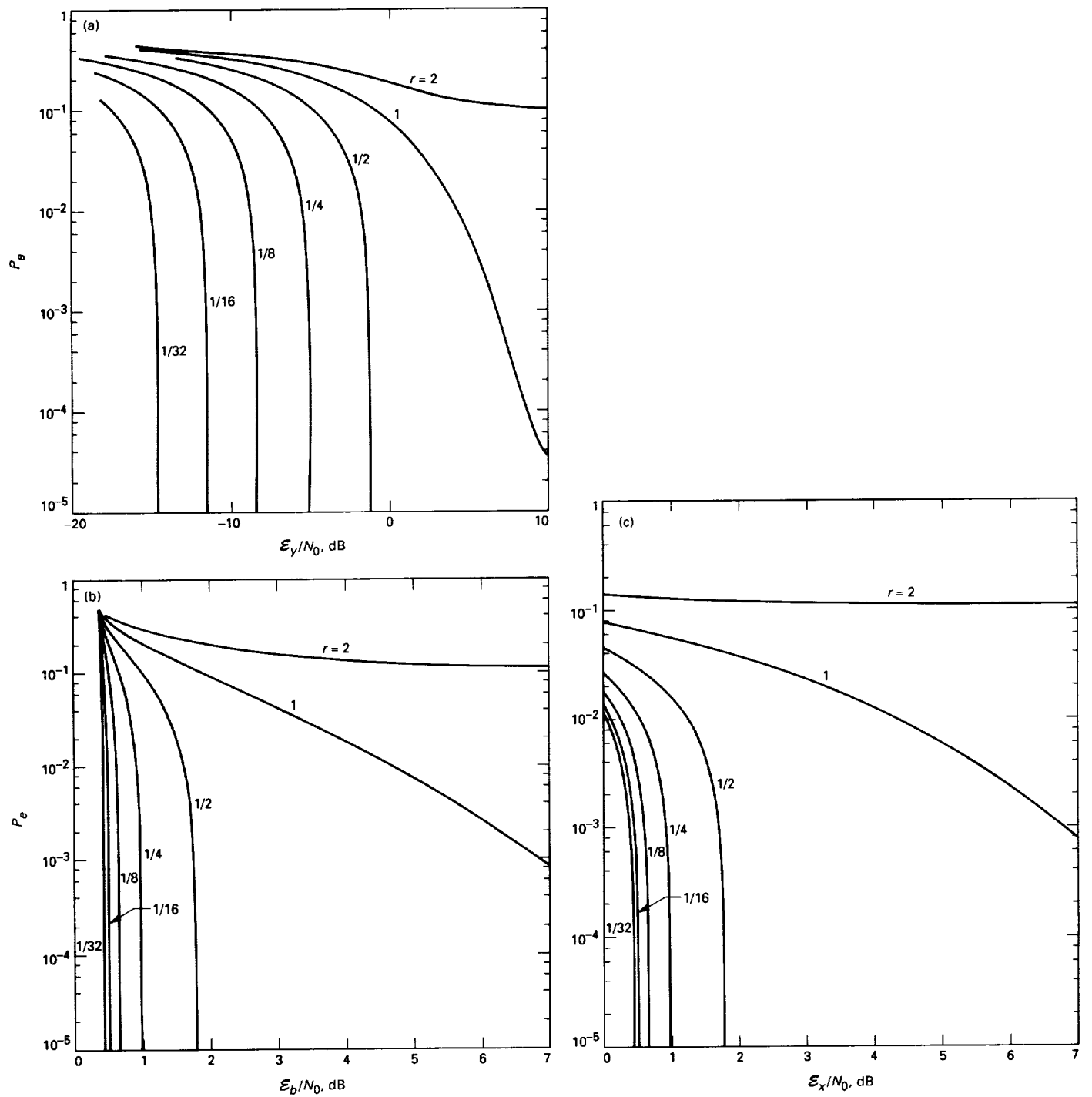


Fig. 7. Source-channel coding limits for a binary symmetric source (BSS) and a hard-quantized Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

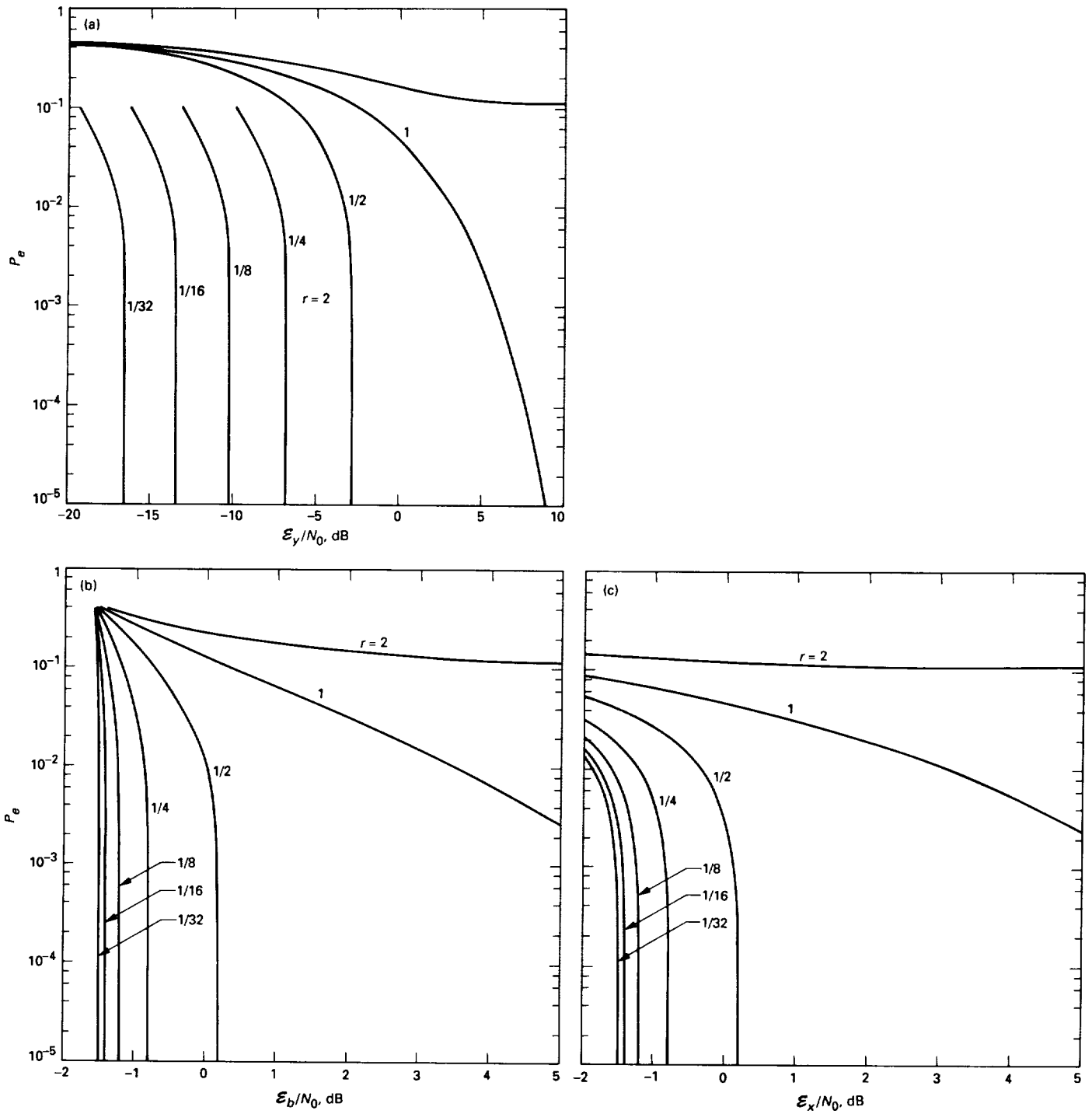


Fig. 8. Source-channel coding limits for a binary symmetric source (BSS) and a binary input Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

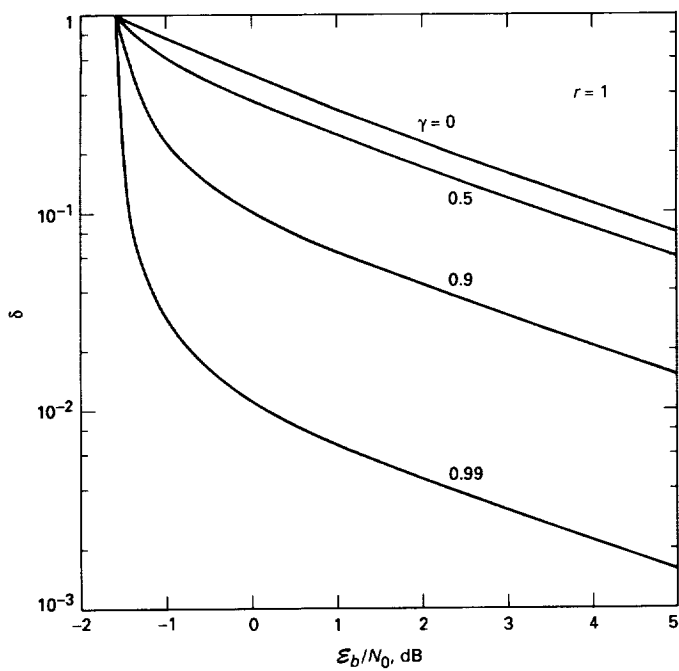


Fig. 9. Source-channel coding limits for a Gauss-Markov source and a Gaussian channel ( $r = 1$ ): distortion versus source-encoded bit signal-to-noise ratio.

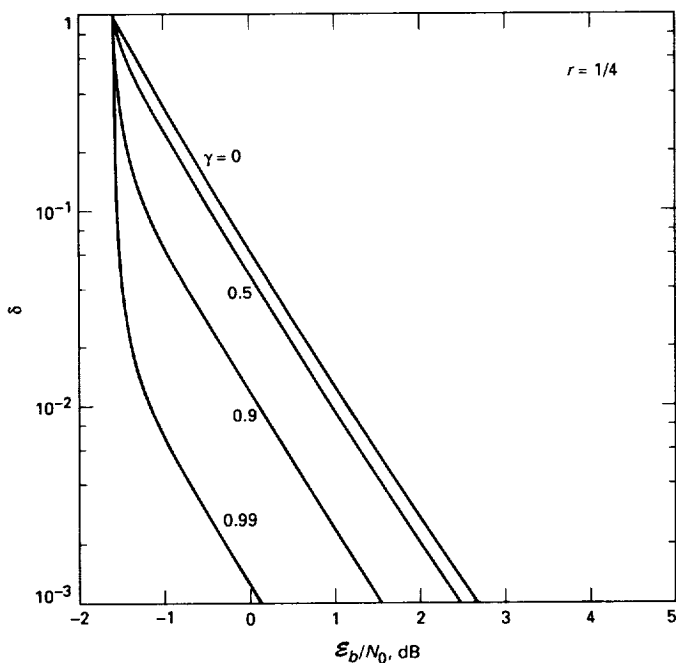


Fig. 10. Source-channel coding limits for a Gauss-Markov source and a Gaussian channel ( $r = 1/4$ ): distortion versus source-encoded bit signal-to-noise ratio.

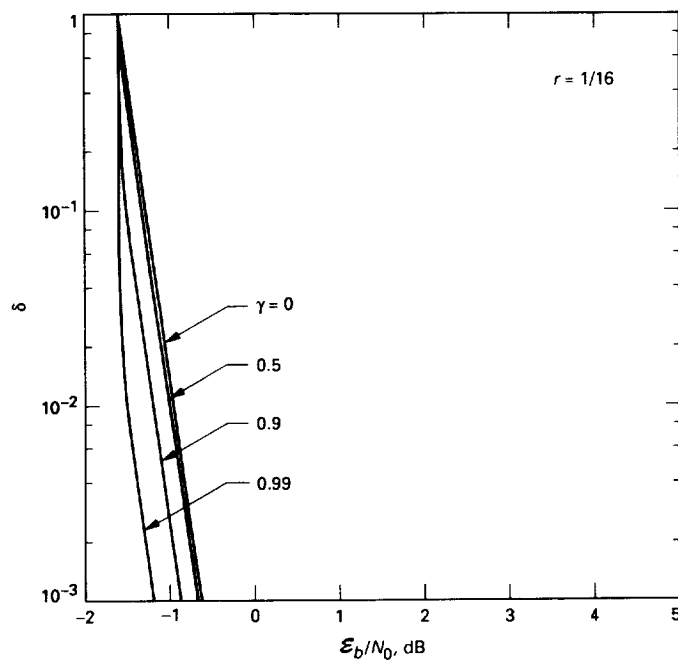


Fig. 11. Source-channel coding limits for a Gauss-Markov source and a Gaussian channel ( $r = 1/16$ ): distortion versus source-encoded bit signal-to-noise ratio.

# A High-Speed Distortionless Predictive Image-Compression Scheme

K.-M. Cheung, P. Smyth, and H. Wang  
Communications Systems Research Section

*This article introduces a high-speed distortionless predictive image-compression scheme that is based on differential pulse code modulation output modeling combined with efficient source-code design. Experimental results show that this scheme achieves compression that is very close to the difference entropy of the source.*

## I. Introduction

Digital data compression is the conversion of a stream of high-rate data (digital or analog) into a stream of relatively low-rate quantized data for communication over a digital communication link or storage in a digital memory. The goals are to reduce the volume of data for transmission over a digital channel and to archive to a digital medium. As the volume of speech and image data in the foreseeable future becomes prohibitively large for many communication links or storage devices, the theory and practice of data compression are receiving increasing attention.

Many data sources contain significant redundancy—symbol distribution, pattern repetition, and positional redundancy. Data compression schemes work by removing redundancy and encoding the new information. They can be broadly categorized into two classes: *lossless* and *lossy*. A lossless coder operating on a digital data source achieves compression without losing any of the digital source information, hence, the decoded image is no different from the original one. A lossy coder reduces the data rate by sacrificing some information that is, or is thought to be, of

little relevance to the user. As a result, the decoded data from a lossy coder show some coding noise, or error, when compared to the original data.

Image data are a typical example of a data source with significant redundancy. Image-compression techniques are used to reduce the volume of raw image data, which is usually too large for many communication links and storage devices. In the past 30 years, there has been a considerable amount of activity in image coding. Much of this research centers around the theme of using elegant compression algorithms to achieve a better compression ratio in the lossless compression scenario, and to achieve a better compression-distortion trade-off in the lossy compression scenario. Due to the recent advances in very large scale integration (VLSI) technology, it is now feasible to implement these algorithms in practical situations, e.g., differential pulse code modulation (DPCM) coding [1], runlength coding [2], the Rice machine [3], and arithmetic coding [4].

DPCM schemes predict the present sample value based on the previous samples and send an encoded version of the difference between the predicted and actual values.

Because of its simplicity, compression efficiency, and ease of implementation, DPCM has become quite common in compressing speech and image data for transmission and storage. In digital processing, the input sequence to the DPCM scheme is generally digitized at the source itself by a fixed number of bits (typically 8 for images). Because predictor values are required to be integer values, the DPCM output sequence will take integer values having a skewed probability distribution. The integer values can then be entropy-coded without distortion using an entropy coder such as the Huffman code [5] or the Rice algorithm [3]. A schematic diagram of a DPCM/entropy coding scheme is shown in Fig. 1.

This article introduces a high-speed distortionless predictive image-compression scheme that is based on DPCM output modeling and efficient source-code design. In general, this scheme is extremely simple to implement and, hence, has potential for high-rate, low-complexity applications. This scheme is particularly useful for spacecraft imaging data, where exact reproduction is a requirement. Other uses include medical image data, where high speed and good quality are both extremely important. The remainder of this article is organized as follows: Section II describes two probabilistic models for the DPCM output integers and Section III outlines the compression algorithm that is based on the models proposed in Section II. Section IV describes the experimental results. Section V discusses the implementation issues. Section VI discusses the entropy and redundancy analysis for this scheme, Section VII addresses the source-model mismatch, and Section VIII gives the concluding remarks.

## II. DPCM Output Modeling

This section describes two probabilistic models for the output samples of a simple DPCM scheme as shown in Fig. 1. To simplify the discussion that follows, a simple predictor consisting of a single delay is chosen. However, the results derived in this article are quite general and apply to DPCM coding schemes with other predictors.

### A. The Modified Laplacian Model

The Laplacian distribution of the form

$$f_L(x) = \frac{1}{\sqrt{2}\sigma} \exp \frac{-\sqrt{2}|x-\mu|}{\sigma} \quad (1)$$

where  $\mu$  denotes the mean and  $\sigma$  denotes the standard deviation, is traditionally used to model the unquantized output samples of a DPCM scheme [6,7]. By requiring the

input sequence and the predictor output samples to be integer values, the DPCM output sequence will take integer values. The probability of the DPCM output integer  $i$ , which is denoted by  $p(i)$ , can be evaluated by integrating from  $i - 0.5$  to  $i + 0.5$ . In most well-behaved speech and image sources,  $\mu$  is very close to zero.

To simplify the discussion that follows, it is assumed that  $\mu = 0$ . To verify the validity of the Laplacian model, four images with 8-bit pixels are selected, and the aforementioned simple DPCM coding scheme is applied to encode the images. For each image, the frequency distribution of the DPCM output samples is tabulated and compared to the corresponding hypothetical distribution given in Eq. (1). The comparisons are shown in Fig. 2. The results were found to be unsatisfactory. The Laplacian distribution was then modified by adjusting the standard deviation  $\sigma$  to  $\lambda\sigma$  such that  $\lambda\sigma = -1/(\sqrt{2}\log(1 - r(0)))$ , and  $r(0)$  is the ratio of the number of zeros in the DPCM output stream to the total number of DPCM output samples. The zero-mean modified Laplacian distribution is given in Eq. (2) as

$$f_{ML}(x) = \frac{1}{\sqrt{2}\lambda\sigma} \exp \frac{-\sqrt{2}|x|}{\lambda\sigma} \quad (2)$$

The introduction of  $\lambda$  corresponds to adjusting the standard deviation  $\sigma$  of the hypothetical distribution by matching the frequency of occurrences of the most probable event in the experiment, namely the integer zero, to the frequency of occurrences of zero predicted by the hypothetical Laplacian model. The comparisons between the experimental frequency distribution and the hypothetical distribution of the images are shown in Fig. 3. The difference entropy, defined as the entropy of the differences between adjacent pixels, is derived from the DPCM output samples and the entropy calculated from the hypothetical modified Laplacian model of the images. The comparisons are shown in Table 1. This modified Laplacian model can be used (1) to simplify the performance analysis of the DPCM scheme and (2) to design an optimal scalar quantizer (e.g., a Lloyd-Max quantizer) for the DPCM output samples.

### B. The Two-Sided Geometric Model

The above continuous modified Laplacian distribution can be shown to provide a good foundation for efficient source-code design. However, this article deals with discrete (integer) DPCM outputs rather than a continuous random variable. The probability distribution of DPCM output integers derived from the modified Laplacian distribution is geometric for all integers except zero. To circum-

vent these problems, a simple single-parameter discrete probability distribution for the DPCM output integers is introduced:

$$p_G(i) = \frac{1-\theta}{1+\theta} \theta^{|i|} \quad \forall i \quad (3)$$

where

$$\theta = \frac{1-r(0)}{1+r(0)}$$

As in the modified Laplacian model introduced in Section II.A, the frequency of occurrences of the most probable event in the experiment, namely the integer zero, matches that predicted by the hypothetical two-sided geometric model. The comparisons between the experimental frequency distribution and the hypothetical distribution of the images are shown in Fig. 4. The two distributions are almost indistinguishable. The difference entropies derived from the hypothetical two-sided geometric model of the images are shown in Table 1. The two-sided geometric model will be used in later sections to derive a simple and efficient lossless coding scheme for the DPCM output samples.

### III. Efficient Coding Based on the Two-Sided Geometric Model

Constructing an optimal prefix code, by using the Huffman algorithm for example, is quite a complex operation in hardware. This section presents efficient techniques to design a near-optimal prefix code to encode images with 8 bits/pixel using the two-sided geometric model introduced in the previous section. In contrast, this code is very simple to implement in hardware. For most well-behaved images,  $\text{frequency}(i) \approx \text{frequency}(-i)$  for  $i = 1, 2, \dots, 255$ . Thus, in order to construct a code for the DPCM output samples, which have values ranging from  $-255$  to  $255$ , one can construct a prefix code for the integer set with values from  $0$  to  $255$ . An additional bit is then appended to each codeword, except the codewords representing  $0$ , to indicate whether integer  $i$  or integer  $-i$  is sent.

Gallager and van Voorhis presented an optimal binary prefix code for the set of geometrically distributed non-negative integers (single-sided geometric) [8]. This technique is adapted and modified as discussed above to encode the two-sided geometrically distributed DPCM output integers. The Gallager and van Voorhis optimal source code for geometrically distributed integer alphabets is a concatenation of a unary and a Huffman code; here it is

named the Gallager-van Voorhis-Huffman (GVH) code. Let  $l$  be the integer satisfying

$$\theta^l + \theta^{l+1} \leq 1 < \theta^l + \theta^{l-1} \quad (4)$$

where

$$\theta = \frac{1-r(0)}{1+r(0)}$$

as defined in Section II.B. It is easy to see that for any  $\theta$ ,  $0 < \theta < 1$ , there is a unique positive integer  $l$  satisfying Eq. (4). Let a non-negative number  $i$  be represented by  $i = lj + r$  where  $j = \lfloor i/l \rfloor$ , the integer part of  $i/l$ , and  $r = [i] \bmod l$ . Gallager and van Voorhis showed that an optimal code for the non-negative integers is the concatenation of a unary code that is used to encode  $j$ , and a Huffman code that is used to encode  $r$ ,  $0 \leq r \leq l-1$ .

Each integer  $r$ ,  $0 \leq r \leq l-1$ , represents an equivalence class modulo  $l$ . Gallager and van Voorhis showed that the integer set  $\{r : 0 \leq r \leq l-1\}$  has a distribution

$$p_r = \frac{1-\theta}{1-\theta^l} \theta^r$$

and the sum of the two least likely letters exceeds the probability of the most likely. The length of the optimal codewords can differ by at most 1. It can be shown that the optimal coding for this integer set is to use codewords of length  $\lfloor \log_2 l \rfloor$  for  $i < 2^{\lfloor \log_2 l \rfloor + 1} - l$ , and codewords of length  $\lfloor \log_2 l \rfloor + 1$  otherwise.

In this article, a simple construction is proposed to generate a Huffman code for the integer set  $\{r : 0 \leq r \leq l-1\}$ . The construction algorithm is as follows:

- (1) Generate the preliminary list  $L$  of  $2^{\lfloor \log_2 l \rfloor}$  binary sequences  $\{00\dots 0, \dots, 11\dots 1\}$ , each of which has length  $\lfloor \log_2 l \rfloor$ .
- (2) Append to each of the last  $l - 2^{\lfloor \log_2 l \rfloor}$  binary sequences in  $L$  either a 0 or a 1 to generate two binary sequences of length  $\lfloor \log_2 l \rfloor + 1$ , and call the new list  $L'$ .

It is not hard to see that  $L'$  has a list of  $l$  prefix-conditioned codewords, with  $2^{\lfloor \log_2 l \rfloor + 1} - l$  codewords of length  $\lfloor \log_2 l \rfloor$ , and the rest of length  $\lfloor \log_2 l \rfloor + 1$ ;  $L'$  is an optimal Huffman code for  $\{r : 0 \leq r \leq l-1\}$ . Thus, each DPCM output integer  $i$  can be efficiently encoded using a concatenation of a unary code, a Huffman code, and a sign bit (except for  $i = 0$ ).

## IV. Experimental Results

A number of images with 8-bit pixels were selected, including both planetary and nonplanetary images. The difference entropy of these images was evaluated, and the images were compressed using the Rice algorithm and the DPCM/GVH code. The difference entropy and the average numbers of bits/pixel used to encode the images using the Rice algorithm and the DPCM/GVH code are shown in Table 2. Experimental results show that in all cases the DPCM/GVH scheme achieves compression that is very close to the difference entropy of the sources, and the compression performances of the Rice algorithm and the DPCM/GVH code are almost indistinguishable.

## V. Some Implementation Issues of the DPCM/GVH Scheme

Like the Rice algorithm [3], the DPCM/GVH code yields a performance (average number of bits per pixel) that approaches the difference entropy of the source. As shown in Sections III and VI, the mathematical theories required to derive this algorithm and to show its efficiency are quite elegant. Its implementation, however, is very simple. This section describes some practical implementation issues of the DPCM/GVH coding scheme.

As discussed in Section III, the decoding process requires finding a unique integer  $l$  such that  $\theta^l + \theta^{l+1} \leq 1 < \theta^l + \theta^{l-1}$ . The range of  $\theta$ , and thus  $r(0)$ , that gives a certain  $l$  can be precalculated, and  $l$  can be found easily by using a table look-up on  $r(0)$ . The maximum value for  $l$  is arbitrary; here it is chosen to be 16. Table 3 gives the ranges of  $r(0)$  for  $l = 1, \dots, 16$ . Note that this table look-up on  $l$ 's is optimally designed for the single-sided geometric distribution [5], and there is no guarantee that it is optimal for the two-sided geometrical distribution described in this article. A look-up table that minimizes the redundancy of the code for the two-sided geometric model can be constructed by direct search, and it is given in Table 4. This is discussed further in Section VI. Once  $l$  is established, the rest of the encoding is trivial. For each DPCM output integer  $i$ ,  $-255 \leq i \leq 255$ , one expresses  $|i| = lj + r$ . Since  $0 \leq |i| \leq 255$ , this operation of finding  $j$  and  $r$  given  $|i|$  can be implemented using a table look-up on 256 entities. Since  $1 \leq l \leq 16$ , there are 16 tables. The total memory size required is  $256 \times 16 \times 2 = 8192$  bytes. Once  $j$  and  $r$  are found,  $j$  is encoded using a unary code (a runlength code with  $j$  zeros followed by a 1), and  $r$ ,  $0 \leq r \leq l-1$ , is encoded using a trivial Huffman code with codewords of lengths  $\lfloor \log_2 l \rfloor$  and  $\lfloor \log_2 l \rfloor + 1$ . The Huffman codes corresponding to different  $l$ 's can be generated easily, and they may be stored in a read-only

memory (ROM). For  $l = 1$ ,  $|i| \bmod l = 0$  for all  $|i|$  and no Huffman code is needed. For  $2 \leq l \leq 16$ , the Huffman codes are stored in a ROM with only  $(2 + 16) \times 15/2 = 135$  entities, each of which is at most 4 bits long. This is shown in Table 5. Finally, a 0 or 1 is appended to the concatenated codeword to indicate the sign of a nonzero  $i$ . Of course, it is obvious from Table 4 that some values of  $l$  in the range  $1 \leq l \leq 16$  will never be used. The total memory size can be reduced further by omitting those entities that correspond to the unused  $l$ 's.

In summary, this scheme involves first counting the number of occurrences of identical consecutive digits in a block (a line or the whole picture) of pixels. A table look-up operation is then performed to find  $l$ , which characterizes the subsequent encoding of the pixels in the block. Each pixel in the block is then encoded with one subtraction operation (difference between the present pixel value and previous pixel value) and two table look-ups.

This scheme can be modified to fit different needs. Instead of encoding the whole image using one  $l$ , one can encode one line at a time. This has the following advantages:

- (1) It makes the encoding more adaptive to local statistics.
- (2) It reduces the buffer size (buffer for one line instead of the whole picture).
- (3) It allows on-the-fly encoding.
- (4) It complies with the fixed-line rate format.

However, a potential disadvantage is that the estimate of  $\theta$  is based on fewer samples and may not be as robust as when the estimate is taken over a large number of samples. As the number of samples used  $n$  is reduced, the possibility of having a mismatch between the model and the actual data is increased.

## VI. The Redundancy of Gallager-van Voorhis-Huffman Codes

It has been demonstrated empirically that the GVH coding scheme yields prefix codes whose average length is very close to the difference entropy of the source, i.e., whose redundancy is quite small. The performance of the GVH codes is now analyzed in more detail, deriving closed-form analytic expressions as a function of  $\theta$  for the redundancy, the mean codelength, and entropy of both single-sided and two-sided integer geometric distributions.

It is both instructive and convenient to first analyze the redundancy for a single-sided integer geometric source, i.e., a source where the letters are labeled  $i$ ,  $0 \leq i \leq \infty$ , where

$$p(i) = (1 - \theta)\theta^i$$

The infinite source is treated (rather than the truncated source where  $0 \leq i \leq 255$ ), since the difference that results from the truncation is negligible. In Gallager and van Voorhis's original paper [8], results were stated for the entropy, mean codelength, and redundancy of the GVH code for this source. Define the entropy of the DPCM source to be  $H(X_1)$  (where  $X_1$  is the discrete random variable corresponding to the single-sided geometric source), the mean codelength of the GVH code for  $X_1$  to be  $\bar{l}_1$ , and the redundancy of said code to be  $r_1$ . Furthermore, let  $l_1(i)$  be the length of the  $i$ th codeword, as determined by the GVH procedure. As described in Section III,

$$l_1(i) = j_i + r_i$$

where

$$j_i = \lfloor i/l \rfloor + 1$$

and

$$\begin{aligned} r_i &= \lfloor \log_2(l) \rfloor, & \text{if } i \bmod l < 2^{\lfloor \log_2(l) \rfloor + 1} - l \\ &= \lfloor \log_2(l) \rfloor + 1, & \text{if } i \bmod l \geq 2^{\lfloor \log_2(l) \rfloor + 1} - l \end{aligned}$$

where  $l$  is a function of  $\theta$  as described in Section III.

Appendix A shows that the entropy of this source can be written as

$$H(X_1) = \frac{H_2(\theta)}{1 - \theta}$$

where  $H_2(p) = -p \log_2(p) - (1 - p) \log_2(1 - p)$  is the binary entropy function. It is also shown that the mean codelength<sup>1</sup> is

$$\bar{l}_1 = \lfloor \log_2(l) \rfloor + 1 + \frac{\theta^k}{1 - \theta^l}$$

where  $k$  is defined as

$$k = 2^{\lfloor \log_2(l) \rfloor + 1} - l$$

Hence, the result is

$$\begin{aligned} r_1 &= \bar{l}_1 - H(X_1) \\ &= \lfloor \log_2(l) \rfloor + 1 + \frac{\theta^k}{1 - \theta^l} - \frac{H_2(\theta)}{1 - \theta} \end{aligned}$$

Since this is an optimal prefix code,  $r_1 \leq 1$ ; and indeed, for most values of  $\theta$ ,  $r_1$  is quite small.

Next, the topic of relevance to DPCM image coding is addressed, namely the coding of the two-sided geometric model introduced in Section II.B, where for each letter  $i$ ,  $-\infty \leq i \leq \infty$ , and

$$p_2(i) = \frac{1 - \theta}{1 + \theta} \theta^{|i|}$$

Once again, for convenience the infinite source alphabet (rather than the truncated case) is treated: the truncated case differs by additive terms on the order of  $\theta^{255}$ , which is vanishingly small for all values of  $\theta$  encountered in practice. It has been found that for planetary images, the difference entropy tends to be in the range of 2.0 to 4.5 bits [3], which corresponds to  $\theta$  in the range from 0.3 to 0.8 bits.

The GVH code for the double-sided model is defined as described earlier, i.e., for  $i \neq 0$ , an extra sign bit is appended to the equivalent codeword for a single-sided source. Define  $\bar{l}_2$  as the mean codelength of such a coding scheme, and  $r_2$  as the resulting redundancy. If  $l_2(i)$  is defined as the length of the codeword assigned to letter  $i$  by this scheme, then

$$\begin{aligned} l_2(i) &= l_1(i) + 1, & i \neq 0 \\ &= l_1(0), & i = 0 \end{aligned}$$

First derive an expression for the mean codelength  $\bar{l}_2$ , as a function of  $\theta$  and  $l$ . Note that  $l$  need not strictly be a function of  $\theta$  as before, i.e., there is no guarantee that the condition described in Section III, namely that  $l$  uniquely satisfies

$$1 + \theta \leq \theta^{-l} < 1 + \frac{1}{\theta}$$

leads to an optimal code for the two-sided source. Hence, one may treat the optimal selection of  $l$  as a separate problem; the derivations to follow are quite valid for any choice of  $l$ . The question of how to find the  $l$ 's that minimize  $r_2$  will be discussed later.

<sup>1</sup> This result is different from that given in [8]. There appears to be a typographical error in their equation for  $\bar{l}_1$ , as they have the term  $\lfloor \log_2(l) \rfloor$  instead of  $\lfloor \log_2(l) \rfloor + 1$ .

$$\begin{aligned}
\bar{l}_2 &= \sum_{i=-\infty}^{i=\infty} p_2(i) l_2(i) \\
&= p_2(0) l_1(0) + \sum_{i=1}^{i=\infty} p_2(i) (l_1(i) + 1) \\
&\quad + \sum_{i=-\infty}^{i=-1} p_2(i) (l_1(|i|) + 1) \\
&= 2 \sum_{i=0}^{\infty} p_2(i) l_1(i) + 1 - p_2(0) - p_2(0) l_2(0)
\end{aligned}$$

But by definition

$$p_2(i) = \frac{p_1(i)}{1 + \theta}$$

which leads to

$$\bar{l}_2 = \frac{2}{1 + \theta} \bar{l}_1 + 1 - p_2(0) - p_2(0) l_2(0)$$

Since, in general

$$l_2(0) = 1 + \lfloor \log_2(l) \rfloor$$

and

$$p_2(0) = \frac{1 - \theta}{1 + \theta}$$

and  $\bar{l}_1$  has been derived earlier, one can write

$$\begin{aligned}
\bar{l}_2 &= \frac{2}{1 + \theta} \left( \lfloor \log_2(l) \rfloor + 1 + \frac{\theta^k}{1 - \theta^l} \right) \\
&\quad + 1 - \frac{1 - \theta}{1 + \theta} - \frac{1 - \theta}{1 + \theta} (1 + \lfloor \log_2(l) \rfloor) \\
&= 1 + \lfloor \log_2(l) \rfloor + \frac{2}{1 + \theta} \left( \theta + \frac{\theta^k}{1 - \theta^l} \right)
\end{aligned}$$

Hence, it is seen that the mean codelength for the two-sided GVH coding scheme is quite similar in form to the one-sided GVH result. Clearly, however, the difference in the two forms may lead to different optimal values of the parameter  $l$  for fixed  $\theta$ , i.e.,  $\bar{l}_1$  and  $\bar{l}_2$  may be minimized by different values of  $l$  over certain ranges of  $\theta$ .

It is shown in Appendix B that the entropy of the two-sided source can be written as

$$\begin{aligned}
H(X_2) &= \sum_{i=-\infty}^{i=\infty} p_2(i) l_2(i) \\
&= \log_2 \left( \frac{1 + \theta}{1 - \theta} \right) - \frac{2\theta \log_2(\theta)}{(1 + \theta)(1 - \theta)}
\end{aligned}$$

where  $X_2$  is the discrete random variable corresponding to the double-sided geometric source. Hence, one can write down a closed-form expression for the redundancy of the coding scheme as a function of  $\theta$  and  $l$ , namely

$$\begin{aligned}
r_2 &= \bar{l}_2 - H(X_2) \\
&= 1 + \lfloor \log_2(l) \rfloor + \frac{2}{1 + \theta} \left( \theta + \frac{\theta^k}{1 - \theta^l} \right) \\
&\quad - \log_2 \left( \frac{1 + \theta}{1 - \theta} \right) + \frac{2\theta \log_2(\theta)}{(1 + \theta)(1 - \theta)}
\end{aligned}$$

One can find the value of  $l$  that minimizes  $r_2$  for a given  $\theta$  by minimizing the terms in  $r_2$  that depend on  $l$ , namely

$$f(l) = \lfloor \log_2(l) \rfloor + \frac{2}{1 + \theta} \left( \frac{\theta^k}{1 - \theta^l} \right)$$

The optimal  $l$  values (over all ranges of  $\theta$  of interest) can be found by direct search. Table 4 shows the ranges of

$$r(0) = \frac{1 - \theta}{1 + \theta}$$

for which each value of  $l$  is optimal,  $1 \leq l \leq 30$ . Note in particular that some values of  $l$  are not used in Table 4 and that the ranges are different from the single-sided case (for small values of  $l$ ) as given by Table 3.

Using the optimal  $l$  values, one can plot  $r_2$  as a function of entropy of the two-sided geometric source, with the results shown in Fig. 5. This characteristic of decreasing redundancy as a function of increasing source entropy is quite general and has been bounded by Smyth [9] for arbitrary distributions. The redundancy of this coding scheme yields only a fraction of a bit of inefficiency for ranges of the source entropy of practical interest. The redundancy of the Huffman code as a function of the source entropy for the two-sided geometric model is also plotted in Fig. 5 for comparison. Clearly, the GVH code is near optimal over

all ranges of the source entropy and, from a practical viewpoint, the redundancy improvement by using a Huffman code over the GVH code is negligible.

## VII. Source-Model Mismatch

An obvious topic of interest for further investigation is the robustness of the coding procedure subject to deviations of the true source from the parametric model, i.e., one would like to know how the statistical variations that can be expected in practice will affect the actual number of bits sent back. The mean codelength for the model will not be exactly equal to the mean codelength transmitted for the data. Of course, as described earlier, excellent empirical performance in the actual encoding of planetary images has been observed, but a more quantitative understanding of the trade-offs involved is desirable. Gilbert [10] looked at techniques for designing codes that avoid having very long codewords and, hence, minimize worst-case performance in the case of source-model mismatch. Longo and Galasso [11] looked at very general cases for the minimum deviations required before a probability distribution is "nearer" another optimal code. However, their bounds generally depend on the smallest differences between source probabilities and, as such, are not of practical use for DPCM models (where some of the

source probabilities are very small). Since neither of these approaches is quite appropriate for this problem, there is considerable room for further work. For flight project applications in particular, it will be very important to understand and quantify the robustness of this type of model-based coding prior to practical implementation.

## VIII. Conclusions

The DPCM/GVH coding scheme described in this article effectively replaces the optimal Huffman coding scheme with a near-optimal look-up table operation for the special case of DPCM noiseless coding of image data. Thus, the scheme is eminently suitable for applications that require very high-rate noiseless coding and/or a very simple hardware implementation of the same. Hence, for example, future NASA flight projects such as the proposed Presidential Lunar-Mars Space Initiative<sup>2</sup> may find this scheme very attractive. In addition, very high-rate, ground-based data communication applications such as video conferencing and medical imaging may also be suitable areas for applying this approach.

---

<sup>2</sup>E. C. Posner, "Implications of the Presidential Lunar and Mars Space Initiative for the DSN," JPL IOM ECP 89-67 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 1990.

## References

- [1] B. N. Oliver, "Efficient Coding," *Bell Syst. Tech. J.*, vol. 31, pp. 724-750, July 1952.
- [2] P. Chen and P. Wintz, "Data Compression for Satellite Images," TR-EE-76-9, School of Electrical Engineering, Purdue University, West Lafayette, Indiana.
- [3] R. Rice, "Some Practical Universal Noiseless Coding Techniques," Part I, JPL Publication 79-22, Jet Propulsion Laboratory, Pasadena, California, March 1979, and Part II, JPL Publication 83-17, Jet Propulsion Laboratory, Pasadena, California, March 1983.
- [4] G. Langdon and J. Rissanen, "Compression of Black-White Images with Arithmetic Coding," *IEEE Trans. Communications*, COM-29, pp. 858-867, June 1981.
- [5] D. A. Huffman, "A Method for the Construction of Minimum Redundancy Codes," *Proc. IRE*, vol. 40, pp. 1098-1101, 1952.
- [6] A. Jain, "Image Data Compression: A Review," *Proc. IEEE*, vol. 69, no. 3, pp. 349-389, March 1981.
- [7] J. Sullivan, "Quantization for MMSE Linear Predictive Image Coding Using Block-Adaptive Gain and Bias," *Proceedings of the International Conference on Digital Signal Processing-87*, Florence, Italy, pp. 507-511, September 7-10, 1987.
- [8] R. Gallager and D. van Voorhis, "Optimal Source Codes for Geometrically Distributed Integer Alphabets," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 228-230, March 1975.
- [9] P. Smyth, "Entropy-Based Bounds on the Redundancy of Prefix Codes," presented at the IEEE International Symposium on Information Theory, San Diego, California, 1990.
- [10] E. N. Gilbert, "Codes Based on Inaccurate Source Probabilities," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 304-314, 1971.
- [11] G. Longo and G. Galasso, "An Application of Informational Divergence to Huffman Codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 36-43, January 1982.

**Table 1. Comparison of the difference entropies of the images, modified Laplacian model, and two-sided geometric model**

Images	Difference entropy	Difference entropy of modified Laplacian model	Difference entropy of two-sided geometric model
Saturn 6	2.945	2.877	3.041
Mercury	4.059	3.976	4.069
Moon	5.504	5.790	5.822
Baboon	6.348	6.387	6.409

**Table 2. Comparisons of the compression performances of the Rice algorithm and the DPCM/GVH code in terms of bit/pixel**

Images	Difference entropy	Rice algorithm	DPCM/GVH code
Saturn 6	2.958	3.030	3.281
USC girl	5.061	5.102	5.272
Lena	5.673	5.812	6.010
Moon	5.514	5.663	5.645
Air scene	6.153	6.268	6.174
Peppers	5.094	5.162	5.215
Baboon	6.352	6.339	6.354
Mercury	4.092	4.234	4.162

**Table 3. Optimal  $l$  values for a single-sided geometric distribution as a function of  $r_0$ , the proportion of zeros in the difference-statistic histogram**

Start of range $r_1(0)$	End of range $r_2(0)$	Optimal $l$
1.000000	0.236094	1
0.236094	0.140251	2
0.140251	0.099505	3
0.099505	0.077586	4
0.077586	0.063264	5
0.063264	0.053741	6
0.053741	0.046572	7
0.046572	0.041124	8
0.041124	0.036807	9
0.036807	0.033058	10
0.033058	0.030397	11
0.030397	0.027749	12
0.027749	0.026167	13
0.026167	0.024066	14
0.024066	0.022495	15
0.022495	0.021450	16
0.021450	0.019888	17
0.019888	0.018849	18
0.018849	0.017812	19
0.017812	0.017294	20
0.017294	0.016260	21
0.016260	0.015744	22
0.015744	0.015228	23
0.015228	0.014199	24
0.014199	0.013685	25
0.013685	0.013171	26
0.013171	0.012658	28
0.012658	0.012146	29

**Table 4. Optimal  $l$  values for a double-sided geometric distribution as a function of  $r_0$ , the proportion of zeros in the difference-statistic histogram**

Start of range $r_1(0)$	End of range $r_2(0)$	Optimal $l$
1.000000	0.296176	1
0.296176	0.140251	2
0.140251	0.126126	3
0.126126	0.077586	4
0.077586	0.063264	5
0.063264	0.055966	7
0.055966	0.041124	8
0.041124	0.036807	9
0.036807	0.033058	10
0.033058	0.030397	11
0.030397	0.027749	12
0.027749	0.026167	15
0.026167	0.021450	16
0.021450	0.019888	17
0.019888	0.018849	18
0.018849	0.017812	19
0.017812	0.017294	20
0.017294	0.016260	21
0.016260	0.015744	22
0.015744	0.015228	23
0.015228	0.014199	24
0.014199	0.013685	25
0.013685	0.013171	26
0.013171	0.012658	28
0.012658	0.012146	29

**Table 5. Codeword tables for  $1 = 1, \dots, 16$**

Codewords	Table number	Codewords	Table number
0 1	2	00 01 100 101 110 111	6
0 10 11	3	00 010 011 100 101 110 111	7
00 01 10 11	4	000 001 010	8
00 01 10 110 111	5		

Table 5 (contd)

Codewords	Table number	Codewords	Table number
011 100 101 110 111	8	0111 1000 1001 1010 1011 1100 1101 1110 1111	13
000 001 010 011 100 101 110 1110 1111	9	000 001 0100 0101 0110 0111 1000 1001 1010 1011 1100 1101 1110 1111	14
000 001 010 011 100 101 1100 1101 1110 1111	10	000 0010 0011 0100 0101 0110 0111 1000 1001 1010 1011 1100 1101 1110 1111	15
000 001 010 011 100 1010 1011 1100 1101 1110 1111	11	0000 0001 0010 0011 0100 0101 0110 0111 1000 1001 1010 1011 1100 1101 1110 1111	16
000 001 010 011 1000 1001 1010 1011 1100 1101 1110 1111	12		
000 001 010 0110	13		

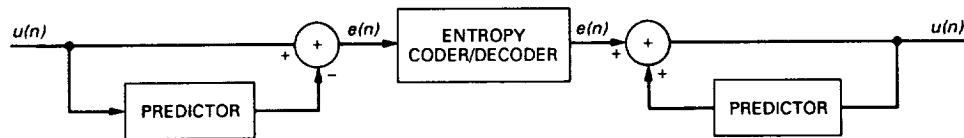


Fig. 1. A DPCM/entropy coder.

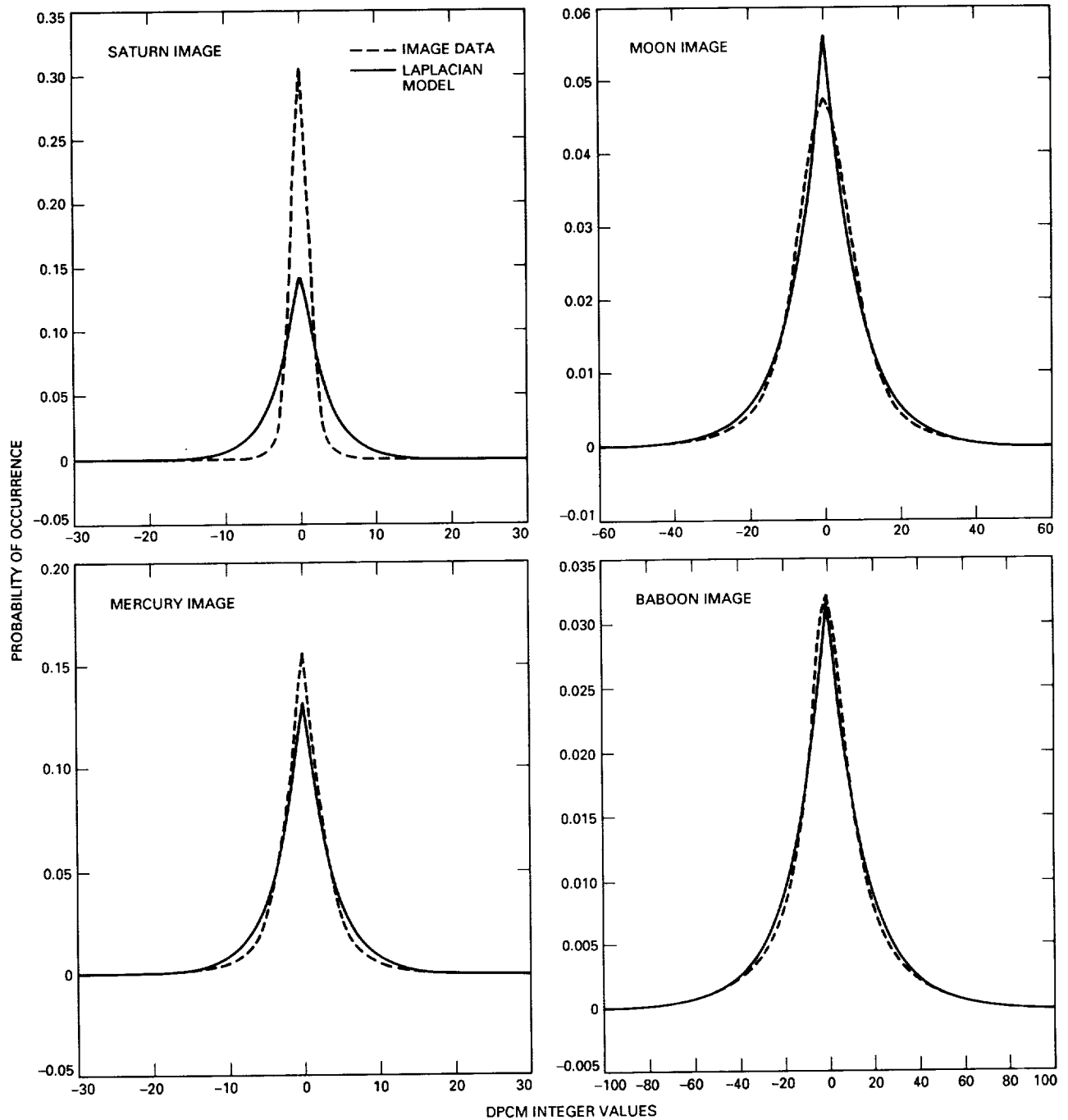
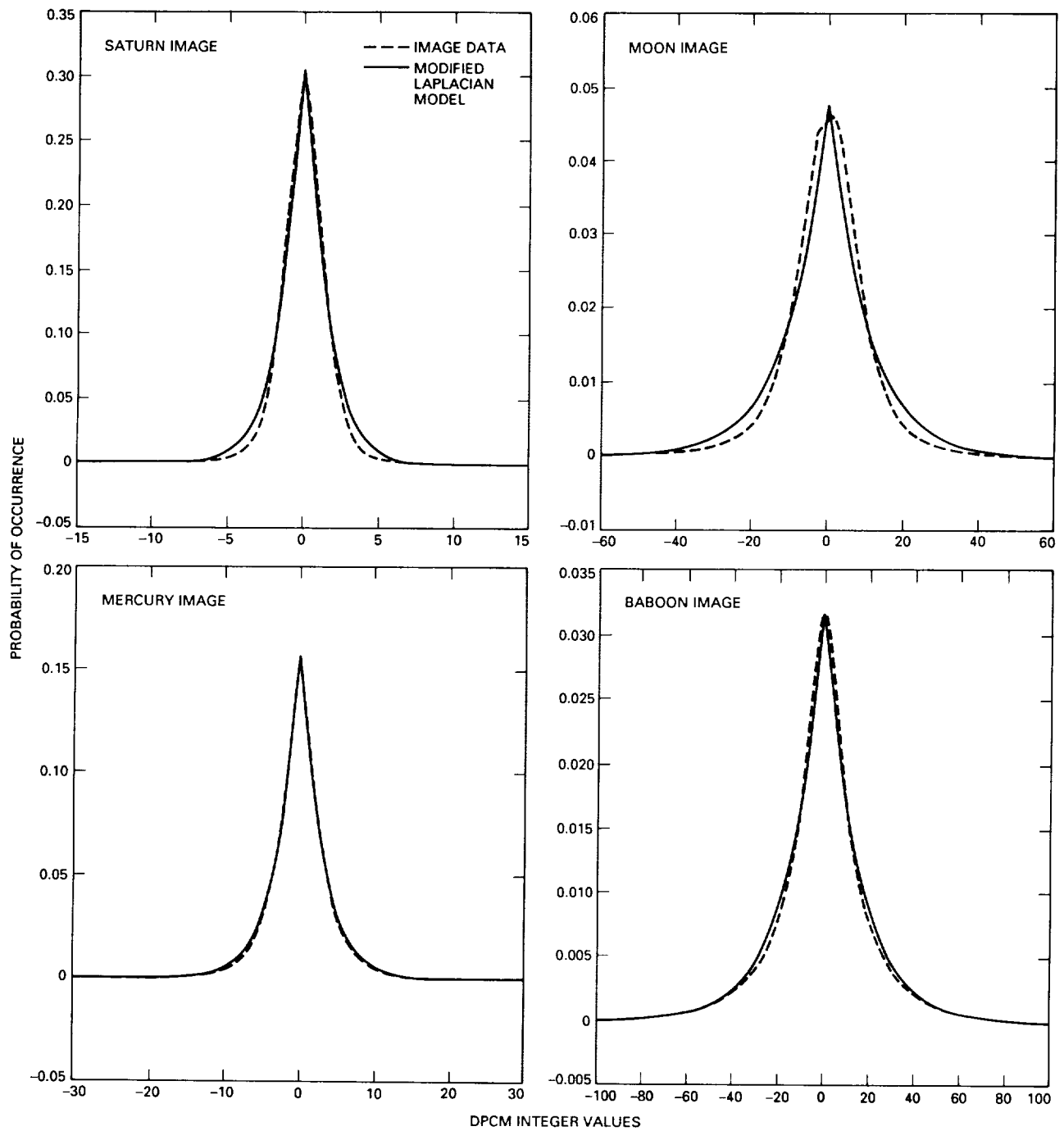


Fig. 2. Comparisons between empirical distributions and simple Laplacian distributions.



**Fig. 3. Comparisons between empirical distributions and modified Laplacian distributions.**

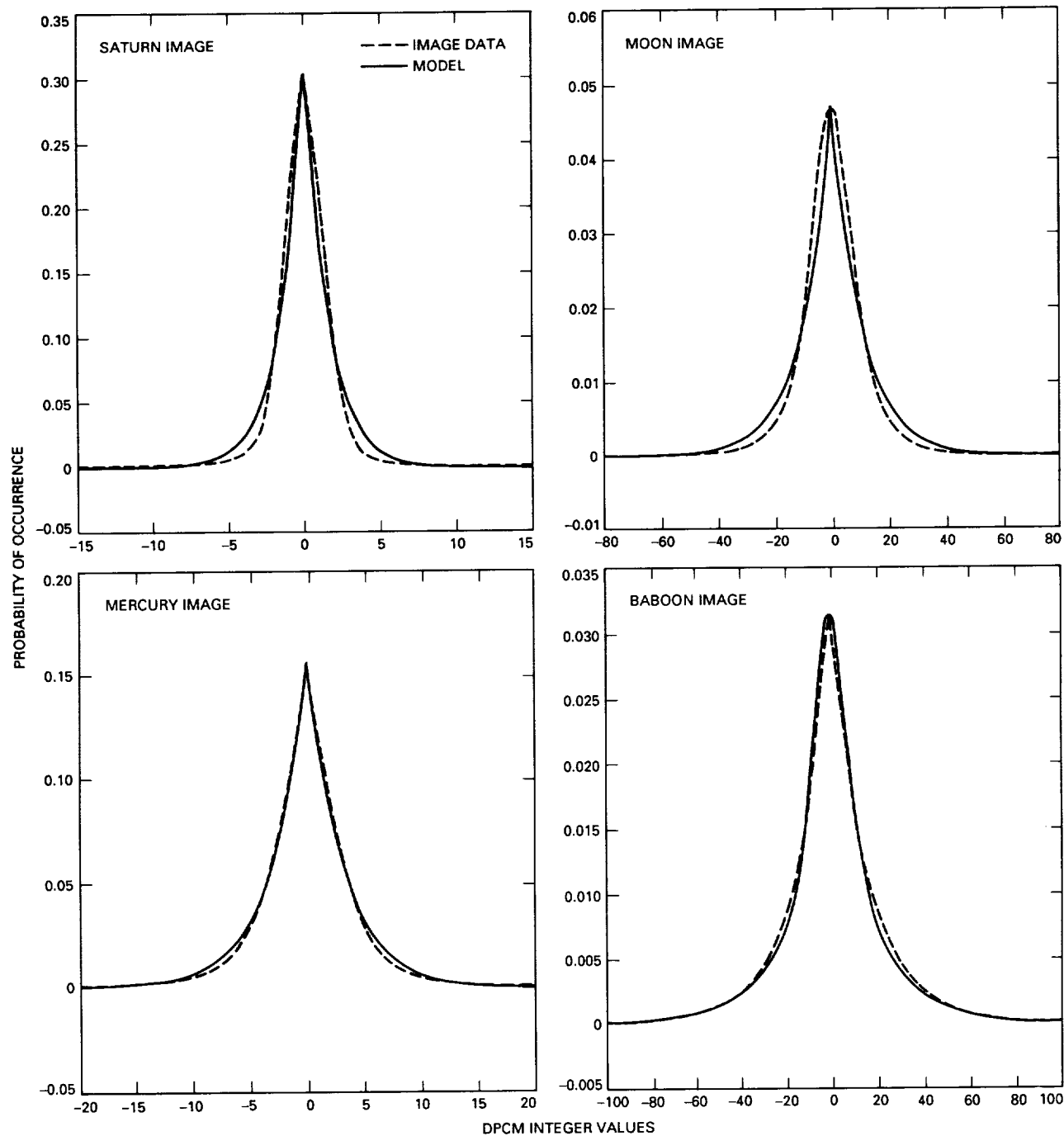
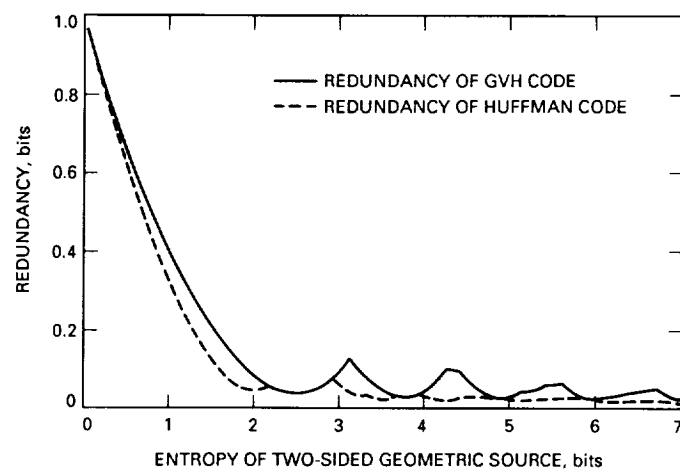


Fig. 4. Comparisons between empirical distributions and two-sided geometric distributions.



**Fig. 5. Redundancy versus entropy for a two-sided geometric source.**

## Appendix A

### Mean Codelength of GVH Code for Single-Sided Geometric Distribution

This Appendix gives an expression for the mean codelength of the GVH code for a single-sided geometric distribution as defined in Section V. One has

$$\begin{aligned}
 \bar{l}_1 &= \sum_{i=0}^{\infty} p_i l_i \\
 &= (1 - \theta) \sum_{i=0}^{\infty} \theta^i l_i \\
 &= (1 - \theta) \left( \sum_{i=0}^{\infty} \theta^i j_i + \sum_{i=0}^{\infty} \theta^i r_i \right) \\
 &= (1 - \theta) \left( \sum_{i=0}^{\infty} \theta^i \left( \left\lfloor \frac{i}{l} \right\rfloor + 1 \right) + \sum_{i=0}^{\infty} \theta^i \left( \lfloor \log_2(l) \rfloor + \delta_i \right) \right)
 \end{aligned}$$

where

$$\begin{aligned}
 \delta_i &= 1 \quad \text{if } i \pmod{l} \geq k \\
 &= 0 \quad \text{otherwise}
 \end{aligned}$$

and

$$k = 2^{\lfloor \log_2(l)+1 \rfloor} - l$$

and so one has

$$\begin{aligned}
 \bar{l}_1 &= (1 - \theta) \left( \sum_{i=0}^{\infty} \theta^i \left\lfloor \frac{i}{l} \right\rfloor + \sum_{i=0}^{\infty} \theta^i \right. \\
 &\quad \left. + \sum_{i=0}^{\infty} \theta^i \lfloor \log_2(l) \rfloor + \sum_{i=0}^{\infty} \theta^i \delta_i \right) \\
 &= S_1 + S_2 + S_3 + S_4
 \end{aligned}$$

Clearly

$$S_2 + S_3 = \frac{1 + \lfloor \log_2(l) \rfloor}{1 - \theta}$$

Looking at the first sum,

$$\begin{aligned}
 S_1 &= \sum_{i=0}^{\infty} \theta^i \left\lfloor \frac{i}{l} \right\rfloor \\
 &= 0(\theta^0 + \theta^1 + \theta^2 + \dots \theta^{l-1}) \\
 &\quad + 1(\theta^l + \theta^{l+1} + \theta^{l+2} + \dots \theta^{2l-1}) \\
 &\quad + 2(\theta^{2l} + \theta^{2l+1} + \theta^{2l+2} + \dots \theta^{3l-1}) \\
 &\quad + \dots \\
 &= \sum_{i=0}^{\infty} i \left( \theta^{il} \sum_{j=0}^{l-1} \theta^j \right) \\
 &= \frac{1 - \theta^l}{1 - \theta} \sum_{i=0}^{\infty} i \left( \theta^l \right)^i
 \end{aligned}$$

Since

$$\sum_{i=0}^{\infty} i \alpha^i = \frac{\alpha}{1 - \alpha^2} \quad 0 < \alpha < 1$$

one gets

$$\begin{aligned}
 S_1 &= \frac{1 - \theta^l}{1 - \theta} \frac{\theta^l}{(1 - \theta)^2} \\
 &= \frac{\theta^l}{(1 - \theta)(1 - \theta^l)}
 \end{aligned}$$

For the fourth sum,  $S_4$ , one gets

$$\begin{aligned}
 S_4 &= \sum_{i=0}^{\infty} \theta^i \delta_i \\
 &= (\theta^k + \theta^{k+1} + \dots + \theta^{l-1}) \\
 &\quad + (\theta^{k+l} + \theta^{k+l+1} + \dots + \theta^{2l-1}) \\
 &\quad + \dots
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=0}^{i=\infty} \theta^{il} \left( \sum_{j=k}^{l-1} \theta^j \right) \\
&= \theta^k \sum_{i=0}^{i=\infty} \theta^{il} \left( \sum_{j=0}^{l-1-k} \theta^j \right) \\
&= \theta^k \left( \frac{1 - \theta^{l-k}}{1 - \theta} \sum_{i=0}^{i=\infty} (\theta^l)^i \right) \\
&= \theta^k \left( \frac{1 - \theta^{l-k}}{1 - \theta} \frac{1}{1 - \theta^l} \right)
\end{aligned}$$

Adding  $S_1 + S_4$ , one gets

$$S_1 + S_4 = \frac{1}{1 - \theta} \frac{\theta^k}{1 - \theta^l}$$

Hence

$$\bar{l}_1 = \lfloor \log_2(l) \rfloor + 1 + \frac{\theta^k}{1 - \theta^l}$$

where  $k$  is defined as

$$k = 2^{\lfloor \log_2(l)+1 \rfloor} - l$$

## Appendix B

### Entropy of a Two-Sided Geometric Source as a Function of $\theta$

This Appendix gives an expression for the entropy of a two-sided geometric source as a function of  $\theta$ . One has

$$\begin{aligned} H(X_2) &= - \sum_{i=-\infty}^{i=\infty} p_i \log_2(p_i) \\ &= - \sum_{i=-\infty}^{i=\infty} \left( \frac{1-\theta}{1+\theta} \right) \theta^{|i|} \log_2 \left( \left( \frac{1-\theta}{1+\theta} \right) \theta^{|i|} \right) \\ &= \left( \frac{1-\theta}{1+\theta} \right) \left( \log_2 \left( \frac{1-\theta}{1+\theta} \right) \left( 1 + 2 \sum_{i=1}^{i=\infty} \theta^i \right) - \log_2(\theta) \left( 2 \sum_{i=1}^{i=\infty} i \theta^i \right) \right) \\ &= \left( \frac{1-\theta}{1+\theta} \right) \left( \log_2 \left( \frac{1-\theta}{1+\theta} \right) \left( 1 + \frac{2\theta}{1-\theta} \right) - \frac{2 \log_2(\theta) \theta}{(1-\theta)^2} \right) \\ &= \log_2 \left( \frac{1-\theta}{1+\theta} \right) - \frac{2\theta \log_2(\theta)}{(1-\theta)(1+\theta)} \end{aligned}$$

50 80  
305350

N91-11985

TDA Progress Report 42-102

August 15, 1990

P-10

# Technical Considerations on Using the Large Nançay Radio Telescope for SETI

S. Gulkis

Space Physics and Astrophysics Section

F. Biraud and J. Heidmann

Observatoire de Paris

Meudon, France

J. Tarter

University of California, Berkeley, California

The SETI Institute, Los Altos, California

*The Nançay decimetric Radio Telescope (NRT) in Nançay, France, is described, and its potential use for SETI (Search for Extraterrestrial Intelligence) observations is discussed. The conclusion reached is that the NRT is well suited for SETI observations because of its large collecting area, its large sky coverage, and its wideband frequency capability. However, a number of improvements are necessary in order to take full advantage of the system in carrying out an efficient SETI program. In particular, system sensitivity should be increased. This can be achieved through a series of improvements to the system, including lowering the ground pickup noise through the use of ground reflectors and more efficient feed design, and by using low-noise amplifier front ends.*

## I. Background

The Nançay Radio Telescope (NRT) is a large astronomical telescope located in Nançay, France, 200 km south of Paris. Figure 1 shows an aerial view of the instrument. The telescope has been in operation since 1964. Administratively, it is operated by the Observatoire de Paris, Département de Radioastronomie Décimétrique.

The configuration of the NRT derives from the concept formulated by J. D. Kraus and used in the construction

of the Ohio State University Radio Telescope (OSURT) in the United States, but the NRT is much larger and has better tracking capability. It is made of two reflecting surfaces, one fixed and one movable. The movable surface is a plane, and can be rotated about a horizontal east-west axis to move the telescope beam in declination. The tiltable surface is 40 m × 200 m (east-west), made up of ten independent elements, each 40 m × 20 m. The fixed reflector is curved in the shape of a portion of a sphere, with a radius of 560 m. Its dimensions are 35 m (vertical) × 300 m (east-west). The spherically shaped reflector facilitates the

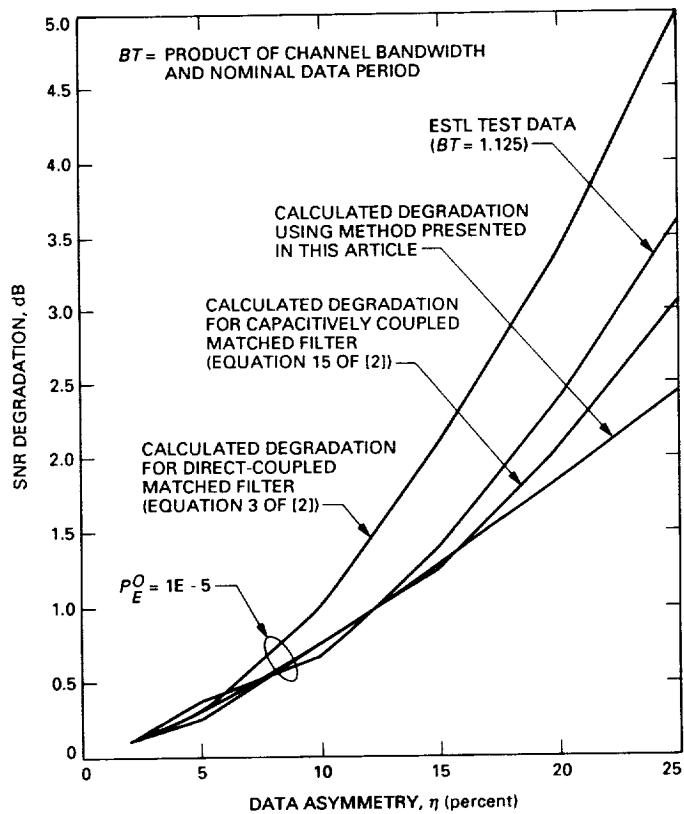


Fig. 8. Performance degradation comparison with ESTL test data for  $BT = 1.125$ .

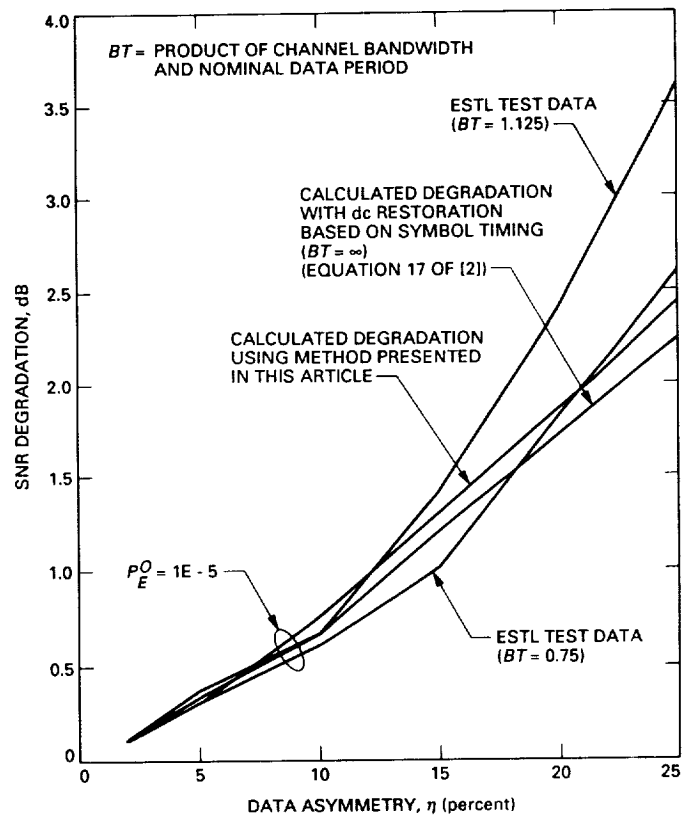


Fig. 9. Performance degradation comparison for  $BT = 1.125$  and  $BT = 0.75$ .

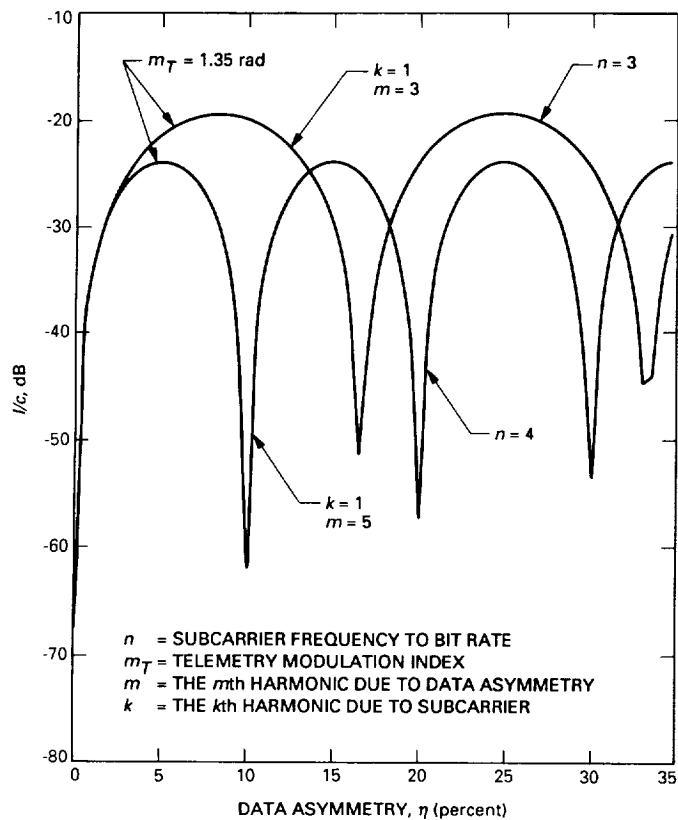


Fig. 6. Interference-to-carrier ratio versus data asymmetry for various values of  $\eta$ .

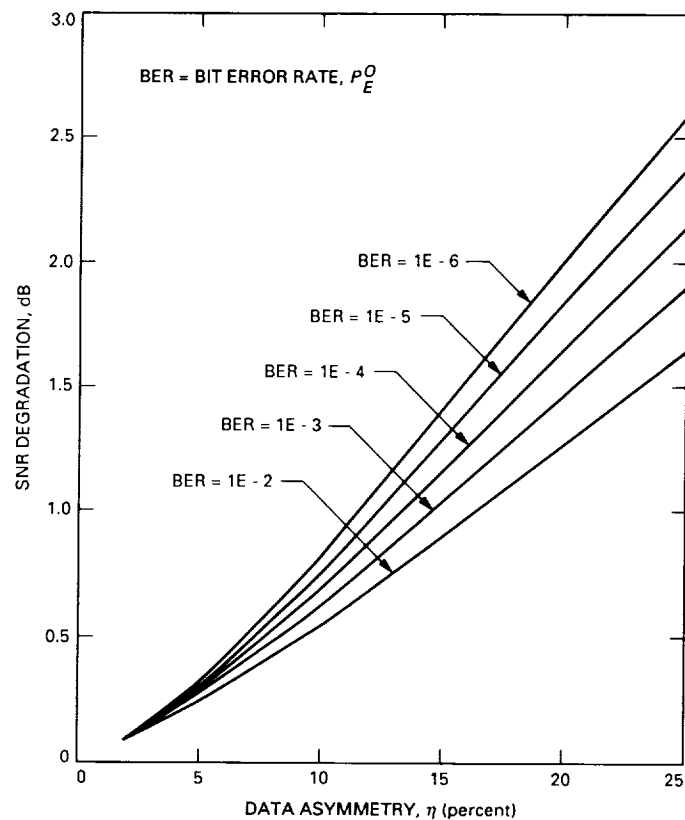


Fig. 7. Telemetry-bit SNR degradation versus data asymmetry.

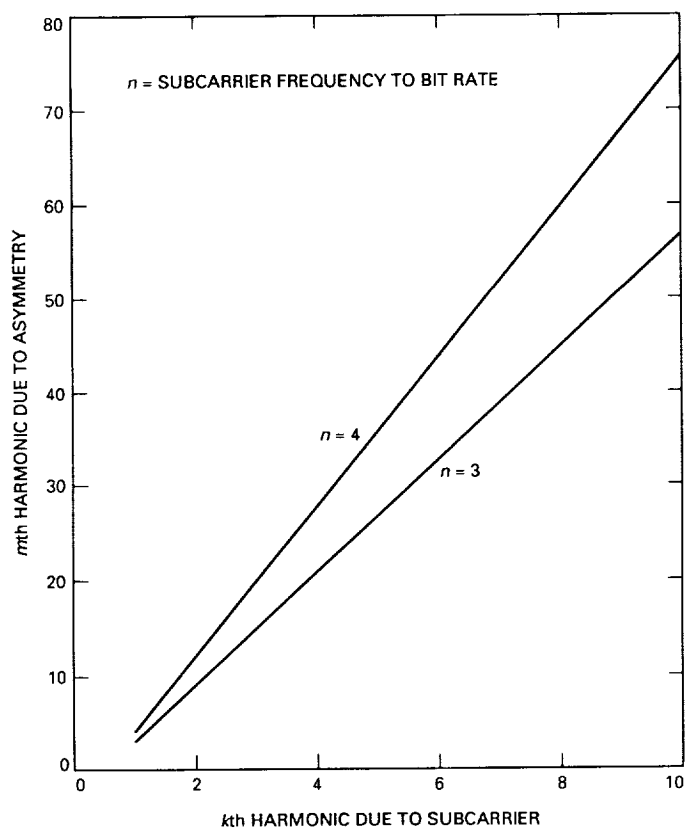


Fig. 4. Functional behavior of the  $m$ th versus  $k$ th harmonic.

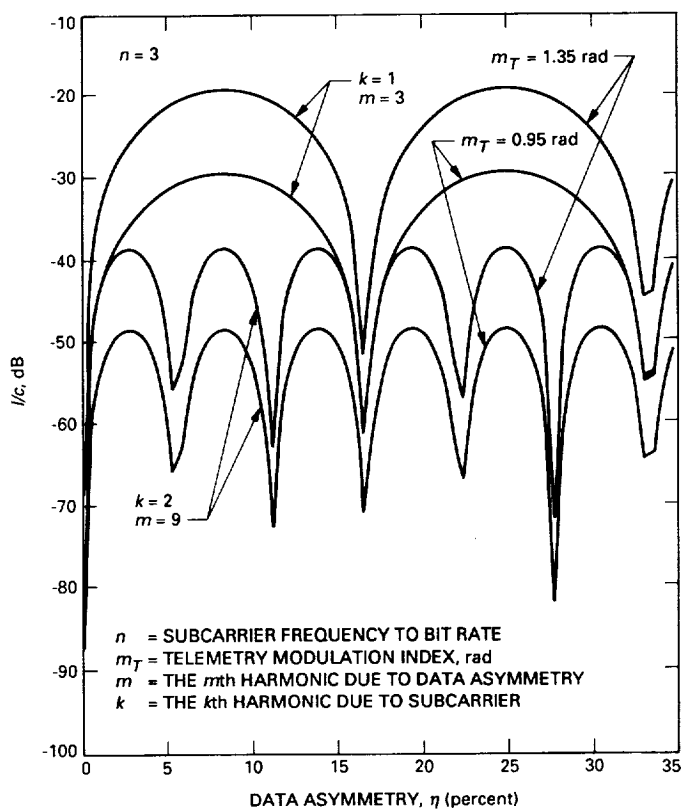


Fig. 5. Interference-to-carrier ratio versus data asymmetry for various values of the telemetry modulation index.

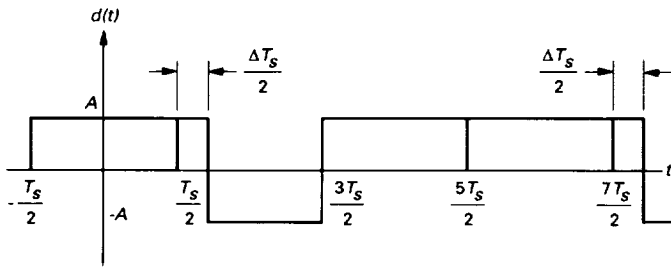


Fig. 1. Asymmetric NRZ data stream using asymmetry model 1 [2].

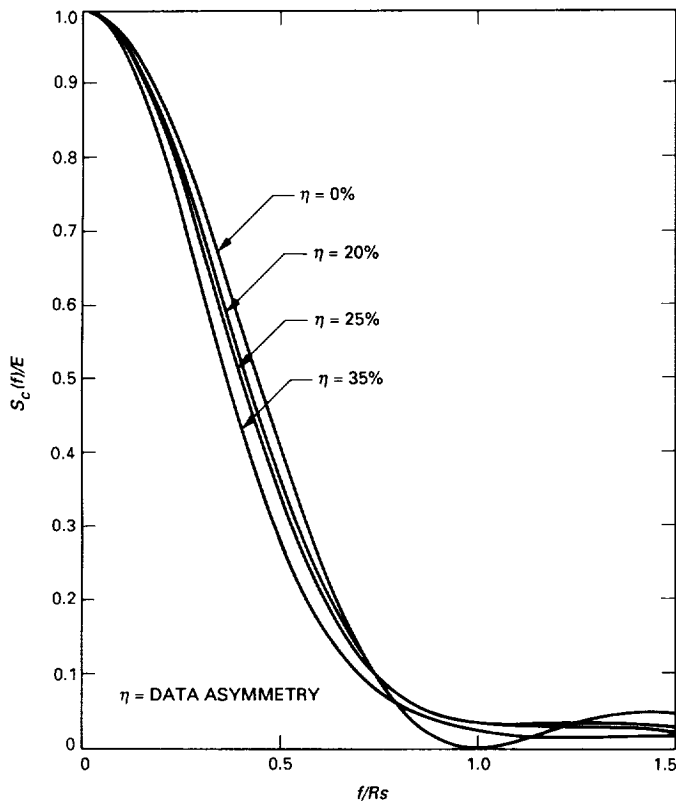


Fig. 2. Normalized power spectra versus data asymmetry.

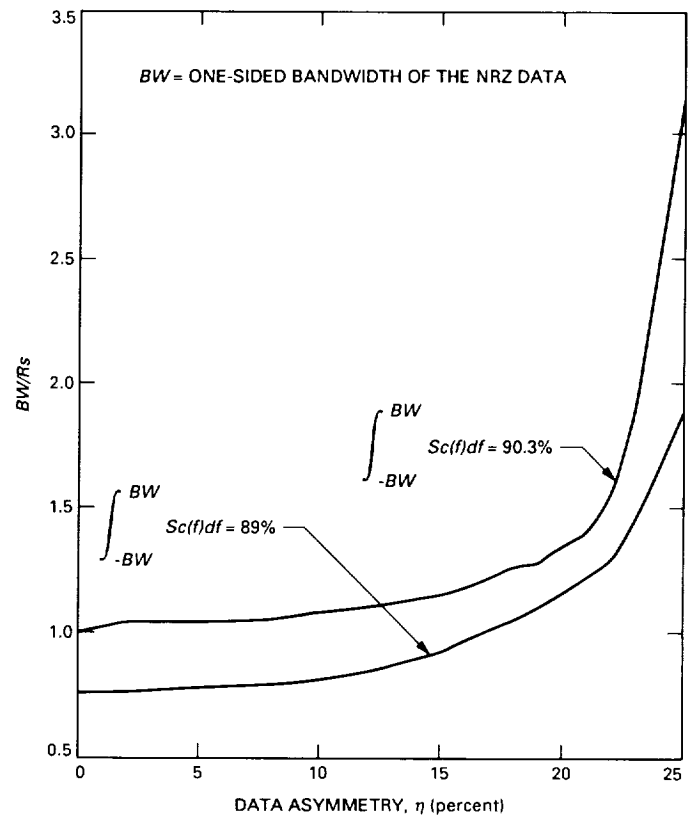


Fig. 3. One-sided bandwidth versus data asymmetry.

**Table 1. Telemetry-bit SNR degradation for uncoded random NRZ data (representing the change in decibels) with  $p = p_t = 1/2$**

$P_E^o$	$\eta$ , percent					
	2	5	10	15	20	25
$10^{-2}$	0.09	0.24	0.53	0.87	1.26	1.69
$10^{-3}$	0.09	0.26	0.60	1.00	1.46	1.94
$10^{-4}$	0.1	0.28	0.66	1.13	1.65	2.19
$10^{-5}$	0.1	0.30	0.73	1.25	1.83	2.42
$10^{-6}$	0.1	0.31	0.79	1.38	2.00	2.63

## Acknowledgments

The author is indebted to Drs. M. K. Simon and F. Davarian for their invaluable comments and useful suggestions during the preliminary review of this work.

## References

- [1] Consultative Committee for Space Data Systems, *Recommendations for Space Data System Standards, Radio Frequency and Modulation Systems, Part 1: Earth Stations and Spacecraft*, CCSDS 401.0-B-1, Blue Book, National Aeronautics and Space Administration, Washington, D.C., January 1987.
- [2] M. K. Simon, K. Tu, and B. H. Batson, "Effects of Data Asymmetry on Shuttle Ku-Band Communications Link Performance," *IEEE Transactions on Communications*, vol. COM-26, no. 11, pp. 1639–1651, November 1978.
- [3] W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, Chapter 1, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.
- [4] *Deep Space Telecommunications Systems Engineering*, Chapter 5, J. Yuen, Editor, New York: Plenum Press, 1983.
- [5] International Radio Consultative Committee, "Protection Criteria and Sharing Considerations Relating to Deep Space Research," *Recommendations and Reports of the CCIR, 1986, XVIIth Plenary Assembly, Volume II: Space Research and Radio Astronomy*, Report 685-2, pp. 278–289, Dubrovnik: CCIR, 1986.
- [6] W. K. Alem, G. K. Huth, and M. K. Simon, *Integrated Source and Channel Encoded Digital Communication System Design Study*, Final Report (R7803-7) under Contract NAS 9-15240, Axiomatix, Marina del Rey, California, March 31, 1978.

Since the NRZ data stream is asymmetric with  $p$  as the probability of occurrence of the  $+A$  symbol and  $(1-p)$  for the  $-A$  symbol, the output of the matched filter depends on the polarity of the symbol over which it is integrating and that of the preceding and succeeding symbols. Thus, the mean value of the matched filter output can be computed by examining eight possible three-symbol sequences. From Appendix A, Fig. 2, of [6], the mean of the matched filter output signal in the presence of data asymmetry can be easily shown to have the following form:

$$\mu = \sqrt{E_b T_s} \left\{ p^3 + (1-p) [p(3-2\eta) + (1-p)^2] \right\} \quad (48)$$

If Eqs. (47) and (48) are substituted into Eq. (46),

$$\Delta_m = 20 \log_{10} \left\{ p^3 + (1-p) [p(3-2\eta) + (1-p)^2] \right\} \quad (49)$$

With equiprobable symbols, i.e.,  $p = 1/2$ , Eq. (49) becomes

$$\Delta_m = 20 \log_{10} \left[ \frac{1}{2} (2 - \eta) \right] \quad (50)$$

The telemetry-bit SNR degradation due to data asymmetry is defined as the sum of the degradations due to undesired spectral components, the spectral distortion and the matched filtering loss. From Eqs. (38), (42), and (49), the bit SNR degradation is given by

$$\Delta = \Delta_u + \Delta_d + \Delta_m \quad (51)$$

Here  $\Delta_u$ ,  $\Delta_d$ , and  $\Delta_m$  are defined in Eqs. (38), (42), and (49), respectively.

Equation (51) is used to calculate the telemetry-bit SNR degradation for the specified amount of data asymmetry with bit-error probability ( $P_E^o$ ) as a parameter. For  $p = p_t = 1/2$ , the results of these calculations are given in Table 1.

A plot based on the results in Table 1 is illustrated in Fig. 7. Here the bit SNR degradation as a function of data asymmetry with probability of error ( $P_E^o$ ) as a parameter is illustrated. This figure clearly shows that the telemetry-bit SNR degradation increases as the data asymmetry increases. This figure also shows that, for a

fixed data asymmetry, the bit SNR degradation increases as the error probability decreases.

In Figs. 8 and 9, the analytical results found in Table 1 are compared with the results presented in [2]. Figure 8 illustrates the bit SNR degradation versus data asymmetry along with the comparable results obtained from [2] corresponding to no dc restoration (direct coupling, Eq. 3 of [2]) and dc restoration by capacitive coupling (Eq. 15 of [2]). In addition, Fig. 8 also shows the test results for  $BT = 1.125$  taken in the ESTL [2] for the sake of comparison, where  $BT$  denotes the product of channel bandwidth and nominal data period. For data asymmetry of less than 20 percent, the results obtained using the simple technique presented in this article are very close to the results obtained from Eq. 15 of [2] for a capacitively coupled matched filter.

The results presented in Table 1 are further compared in Fig. 9 with the results obtained in [2] for a specific implementation of the symbol synchronizer with dc restoration based on symbol timing (Eq. 17 of [2]). The test results taken at ESTL for  $BT = 1.125$  and  $0.75$  are also shown in this figure for the sake of comparison. Again, these results are in good agreement. However, it is to be noted that the analytical results predicted by Eq. (51) were derived for a special case where the one-sided channel bandwidth  $BW$  is equal to the data bit rate  $R_s$ , i.e.,  $BT = 1$ .

## V. Conclusions

The impact of the data asymmetry on the performance of a space telemetry system has been investigated based on the derived spectral density of an asymmetric NRZ data stream. The derived power spectrum is also used to investigate how rapidly the data bandwidth increases as the amount of data asymmetry increases. It was shown that the data bandwidth increases exponentially as the amount of data asymmetry increases. An analytical model to predict the threshold levels of undesired spectral components that fall into the carrier-tracking-loop bandwidth was derived, and a simple model to predict the telemetry-bit SNR degradation due to data asymmetry was developed. It was shown that, for a specific operating condition, the carrier tracking loop does not drop lock at any value of  $\eta$ , the data asymmetry, and that the simple model developed in this article can be used to accurately predict the bit SNR degradation.

sources of degradation that the data asymmetry created, namely:

- (1) Undesired discrete spectral components
- (2) Continuous spectral distortion
- (3) Matched filtering loss

The undesired discrete spectral components created by data asymmetry can cause potential interference to the telemetry data channel. If one denotes  $\Delta_u$  as the bit SNR degradation due to the presence of the undesired spectral components, then it is straightforward to show that

$$\Delta_u = 10 \log_{10} \left[ 1 + \frac{1}{N_0(BW)} \int_0^{BW} S_I(f) df \right] \quad (38)$$

where  $N_0$  is the one-sided noise spectrum density,  $BW$  is the one-sided data bandwidth, and  $S_I(f)$  is the interference power spectrum density.

The above equation assumes that the total noise spectral density seen by the telemetry data channel is the thermal noise plus the interference spectral density. If the one-sided bandwidth is equal to the bit rate, i.e.,  $BW = R_s$ , then, from Eq. (18),

$$S_I(f) = S_{dc}(f) + \frac{2A^2 P_t^2}{\pi^2} C(1, p, \eta) \delta(f - R_s) \quad (39)$$

where  $S_{dc}(f)$  is defined in Eq. (18), and  $C_t(1, p, \eta)$  is given in Eq. (24) with  $m = 1$ .

For  $p = P_t = 1/2$ ,  $S_{dc}(f) = S_{dc}^o(f)$ , see Eq. (30). Thus, from Eq. (30), Eq. (39) can be rewritten

$$S_I(f) = A^2 (0.5\eta)^2 \delta(f) + \frac{A^2}{2\pi^2} C(1, 1/2, \eta) \delta(f - R_s) \quad (40)$$

In this case,  $C(1, 1/2, \eta) = (1/4) \sin^2(2\pi\eta)$  is given in Eq. (31) with  $m = 1$ .

Therefore, for  $p = p_t = 1/2$ , from Eq. (40), Eq. (38) can be written as

$$\Delta_u = 10 \log_{10} \left[ 1 + \frac{E_b}{N_0} \left( (0.5\eta)^2 + \frac{C(1, 1/2, \eta)}{2\pi^2} \right) \right] \quad (41)$$

Here,  $E_b = A^2 T_s$  = bit energy. As an example for Category B, the bit energy is found to be:  $E_b = (A_T^2 T_s) \sin^2(m_T)$ , see Eq. (34).

Equation (41) assumes that all the harmonics with frequency components greater than  $R_s$ , i.e.,  $f > R_s$ , have been filtered out by the predetection filter (assumed to be of the brick-wall type).

Figure 2 shows that the continuous component of the telemetry spectral density is distorted by data asymmetry. Spectral distortion can cause further degradation in bit SNR. If  $\Delta_d$  is the bit SNR degradation due to spectral distortion, then  $\Delta_d$  can be defined as

$$\Delta_d = 10 \log_{10} \left[ \frac{a_1}{a_2} \right] \quad (42)$$

where

$$a_1 = \int_{-R_s}^{R_s} S_{cn}(f) df \quad (43)$$

Here,  $S_{cn}(f) = S_c(f)/E_b$ , with  $S_c(f)$  as defined in Eq. (18). For  $p = p_t = 1/2$ ,  $S_c(f) = S_c^o(f)$ . Here,  $S_c^o(f)$  is given in Eq. (30). Then

$$a_2 = \int_{-R_s}^{R_s} S_{ud}(f) df \quad (44)$$

Here,  $S_{ud}(f)$  is the undistorted normalized (by  $E_b$ ) spectral density, which is given by [3]

$$S_{ud}(f) = \left[ \frac{\sin(\pi f/R_s)}{(\pi f/R_s)} \right]^2 \quad (45)$$

It is well known that for binary NRZ data signals, the matched filter is an integrate and dump filter. Since the binary NRZ data stream is distorted by the rising and falling voltage transitions, the integrate and dump filter is no longer the matched filter for detecting the binary NRZ data signal. If  $\Delta_m$  is the bit SNR degradation due to the matched filter, then  $\Delta_m$  can be defined as

$$\Delta_m = 20 \log_{10} \left[ \frac{\mu}{\mu_0} \right] \quad (46)$$

Here the mean value of the matched filter output signal of the undistorted signal is represented by  $\mu$ , and the mean value of the matched filter output signal of the distorted signal is represented by

$$\mu_0 = \sqrt{E_b T_s} \quad (47)$$

is straightforward to show that the power spectral density of the deep-space telemetry signal is equal to

$$S_T(f) = A_T^2 \cos^2(m_T) \delta(f - f_c) + \frac{4}{\pi^2} A_T^2 \sin^2(m_T) \times \sum_{k=1}^{\infty} \left[ \frac{S_d(f - f_c - (2k-1)f_{sc}) + S_d(f - f_c + (2k-1)f_{sc})}{(2k-1)^2} \right] \quad (34)$$

Here,  $S_d(f)$  is given in Eq. (18) with  $A = 1$ .

The index value  $k$ , see Eq. (34), denotes the  $k$ th harmonic due to the subcarrier, and  $m$ , see Eq. (18), denotes the  $m$ th harmonic due to data asymmetry. When mixing the  $m$ th harmonic with the  $k$ th harmonic, a "third harmonic" is produced. This third harmonic can have a spectral component that falls into the carrier-tracking-loop bandwidth. It is desirable to locate this third harmonic by examining all possible combinations of the index values  $k$  and  $m$ . From Eqs. (18) and (34), it is easy to show that when

$$m = 2nk - n \quad (35)$$

for  $k$  and  $n$  are integers. The harmonics  $S_h(f)$ , from Eq. (18), when mixed with the subcarrier signal, will produce spectral components that fall on the top of the carrier spectral component. Here,  $n$  is the ratio (assumed to be an integer) of the subcarrier frequency-to-bit rate, i.e.,  $n = f_{sc}/R_s$ . The number of subcarrier cycles per symbol is chosen so that the interaction between data sidebands and the residual RF carrier is minimum. In practice,  $f_{sc}$  is usually chosen to be three times the data rate, i.e.,  $n = 3$ .

The plot of Eq. (35) is shown in Fig. 4, which illustrates the functional behavior of the  $k$ th harmonic due to subcarrier versus  $m$ th harmonic due to data asymmetry. From this figure, for a given  $k$ , one can determine the corresponding  $m$ th harmonic that, when mixed with the  $k$ th harmonic, will produce a spectral component that falls on the top of the carrier spectral component causing potential interference to the carrier tracking loop. Knowing the values of  $m$  and  $k$ , the interference-to-carrier power ratio  $I/C$  can be calculated. From Eqs. (18) and (34), it is easy to show that

$$I/C = \frac{8p_i^2 \tan^2(m_T) C(m, p, \eta)}{\pi^4 m^2 (2k-1)^2} \quad (36)$$

For  $m = 2nk - n$ ,  $k$  and  $n$  are integers.

For  $p = p_i = 1/2$ , Eq. (34) becomes

$$I/C = \frac{4 \tan^2(m_T) C(m, 1/2, \eta)}{\pi^4 m^2 (2k-1)^2} \quad (37)$$

For  $m = 2nk - n$ ,  $k$  and  $n$  are integers, and where  $C(m, 1/2, \eta)$  is given in Eq. (31).

The plot of Eq. (37) is shown in Figs. 5 and 6 for various values of  $n$ ,  $k$ , and  $m$ . As expected, Fig. 5 shows that the  $I/C$  increases as the telemetry modulation index increases. This occurs because less power is allocated to the carrier component as the telemetry modulation index increases. Furthermore, Fig. 5 shows that the  $I/C$  decreases as both  $k$  and  $m$  increase. Figure 6 shows that the  $I/C$  decreases as  $n$ , the subcarrier frequency-to-bit rate, increases.

Figures 5 and 6 can be used to determine the critical amount of data asymmetry that will produce the threshold level of  $I/C$ . As an example, from Fig. 5, for  $m_T = 1.35$  rads,  $n = 3$ ,  $k = 1$ , and  $m = 3$ , one finds that  $I/C = -19.46$  dB for  $\eta = 8.27$  percent. However, for a deep-space mission, the maximum allowable CW interference for the carrier tracking is  $(I/C)_{\max} = -15$  dB [5]. Thus, for this particular case, the undesired spectral component generated by data asymmetry does not exceed the maximum allowable interference when the data asymmetry  $\eta$  is equal to 8.27 percent. Figure 5 shows that, for  $m_T = 1.35$  rads, 0.95 rads,  $n = 3$ ,  $k = 1, 2$ , and  $m = 3, 5$ , the interference-to-carrier power ratio does not exceed the maximum allowable interference, i.e.,  $I/C < (I/C)_{\max}$ , for any value of  $\eta$ .

#### IV. Telemetry-Bit SNR Degradation Due to Data Asymmetry

To quantitatively determine the telemetry-bit SNR degradation due to data asymmetry, one must define all the degradation sources created by the data asymmetry. In this section, it is assumed that the telemetry BER degradation due to imperfect bit synchronization is negligible (only on the order of a few tenths of a dB) [2]. Furthermore, the telemetry BER degradation due to imperfect subcarrier tracking is also ignored. This is the case because the subcarrier frequency is usually much higher than the data symbol rate; hence, the subcarrier tracking loop will not be affected by the data asymmetry. This is particularly true for a specific implementation of the subcarrier tracking loop described in [4]. Thus, there remain three major

$$a_5(p, p_t) = 0.5p_t(1 - p_t)(1 - 2p) \quad (23)$$

$$C(m, p, \eta) = \sin^2(m\pi\eta) \left[ \cos^2(m\pi\eta) - (1 - 2p)^2 \sin^2(m\pi\eta) \right] \quad (24)$$

$$\eta = \frac{\Delta}{2} \quad (25)$$

Equation (18) shows that the spectrum of the asymmetric NRZ data stream is a function of the probability of occurrence  $p$  of the positive symbol, the transition density  $p_t$ , and the amount of data asymmetry  $\eta$ . However, the power spectral density can be expressed in terms of the transition density by solving Eq. (17) for  $p$  and substituting the result in Eq. (18). Note that Eq. (18) is valid only for  $p_t \leq 1/2$ .

It is also interesting to evaluate the power spectral density  $S_d(f)$  for the case of random data with equiprobable symbols, i.e.,  $p = p_t = 1/2$ . For this special case, from Eq. (19) through (23),

$$a_1(1/2) = 3/8 \quad (26)$$

$$a_2(1/2, 1/2, \eta) = (5/8) \cos^2(\pi f T_s \eta) \quad (27)$$

$$a_3(1/2, 1/2, \eta) = (3/8) \cos^2(\pi f T_s) + (1/8) \cos^2(\pi f T_s \eta) \quad (28)$$

$$a_4(1/2, 1/2, \eta) = a_5(1/2, 1/2) = 0 \quad (29)$$

Where the terms from Eqs. (26) through (29) are substituted into Eq. (18), the power spectral density for the asymmetric NRZ random data stream with equiprobable symbols is

$$S_d^\circ(f) = (A^2 T_s / 8) \left[ \frac{\sin^2(\pi f T_s)}{(\pi f T_s)^2} \right] \left[ 3 + 5 \cos^2(\pi f T_s \eta) \right] + (A^2 T_s / 8) \underbrace{\left[ \frac{\sin^2(2\pi f T_s \eta)}{(\pi f T_s)^2} \right] \left[ 3 \cos^2(\pi f T_s) + \cos^2(2\pi f T_s \eta) \right]}_{\text{continuous spectrum, } S_c^\circ(f)} + \underbrace{A^2 (0.5\eta)^2 \delta(f)}_{\text{dc component, } S_{dc}^\circ(f)}$$

$$+ \underbrace{\frac{A^2}{2\pi^2} \sum_{m=1}^{\infty} \frac{1}{m^2} C(m, 1/2, \eta) \delta(f - mR_s)}_{\text{harmonics, } S_h^\circ(f)} \quad (30)$$

Here, from Eq. (24),  $C(m, 1/2, \eta)$  is

$$C(m, 1/2, \eta) = (1/4) \sin^2(2m\pi\eta) \quad (31)$$

The continuous spectrum  $S_c^\circ(f)$  of Eq. (30) is plotted in Fig. 2, which illustrates the normalized data power spectra (i.e.,  $S_c^\circ(f)/E$ , here  $E = A^2 T_s$ ) for several data asymmetry  $\eta$ . If  $BW$  is defined as the one-sided bandwidth of the NRZ data stream, then the total data power that lies in the frequency domain  $(-BW, BW)$  can be determined. Figure 3 shows the functional dependence of  $BW$  on the data asymmetry for 90.3 percent and 89 percent of the total data power containment. As expected, for a fixed data rate, the one-sided bandwidth  $BW$  increases as  $\eta$ , the data asymmetry, increases.

### III. The Effects of Data Asymmetry on the Carrier Tracking Loop

The space telemetry signal recommended by the CCSDS can be expressed mathematically as [1]

$$s_T(t) = \sqrt{2} A_T \sin[\omega_c t + m_T d(t) P(t)] \quad (32)$$

where  $A_T$  is the transmitted rms voltage,  $\omega_c = 2\pi f_c$  is the angular carrier center frequency in rads/sec,  $m_T$  is the telemetry modulation index in rads,  $d(t)$  is the data waveform generated by the binary ( $\pm 1$ ) NRZ data sequence, and  $P(t)$  is the subcarrier waveform.

The most interesting subcarrier waveforms in space applications are square-wave for deep-space missions (Category B) and sine-wave for near-Earth missions (Category A) [1]. For example, the Category B subcarrier waveform can be represented by the Fourier series expansion

$$P(t) = \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{\cos[(2k-1)2\pi f_{sc} t]}{2k-1} \quad (33)$$

Here,  $f_{sc}$  is the subcarrier frequency in Hz.

If the data sequence  $d(t)$  is asymmetric as shown in Fig. 1 (with equiprobable symbols and  $A = 1$ ), then the power spectral density  $S_d(f)$  of this asymmetric NRZ data stream is found in Eq. (18). From Eqs. (32) and (33), it

$$g_3(t) = \begin{cases} +A & -(T_s/2) < t \leq (T_s/2) \\ 0 & \text{elsewhere} \end{cases} \quad (4)$$

$$g_4(t) = \begin{cases} -A & -(T_s/2) < t \leq (T_s/2) \\ 0 & \text{elsewhere} \end{cases} \quad (5)$$

The Fourier transforms  $G_1(f)$ ,  $G_2(f)$ ,  $G_3(f)$ , and  $G_4(f)$  of the symbols  $g_1(t)$ ,  $g_2(t)$ ,  $g_3(t)$ , and  $g_4(t)$ , respectively, can be shown to have the following forms:

$$G_1(f) = \frac{Ae^{-(j\pi f T_s \Delta)/2}}{\pi f} \sin(\pi f T_s (1 + \Delta/2)) \quad (6)$$

$$G_2(f) = -\frac{Ae^{(j\pi f T_s \Delta)/2}}{\pi f} \sin(\pi f T_s (1 - \Delta/2)) \quad (7)$$

$$G_3(f) = \frac{A \sin(\pi f T_s)}{\pi f} \quad (8)$$

$$G_4(f) = -G_3(f) \quad (9)$$

From Eqs. (6), (7), (8), and (9),

$$G_1(0) = \lim_{f \rightarrow 0} G_1(f) = AT_s(1 + \Delta/2) \quad (10)$$

$$G_2(0) = \lim_{f \rightarrow 0} G_2(f) = -AT_s(1 - \Delta/2) \quad (11)$$

$$G_3(0) = -G_4(0) = \lim_{f \rightarrow 0} G_3(f) = AT_s \quad (12)$$

If  $p$  is the probability of transmitting a positive pulse, i.e.,  $p = \Pr\{g_i(t) = +A\}$ , and  $p_t$  the transition density, i.e.,  $p_t = \Pr\{g_i(t) \neq g_{i+1}(t)\}$ , then  $p_i$ , for  $i = 1, 2, 3$ , and 4 is given by

$$p_1 = \Pr\{g_i(t) = g_1(t)\} = pp_t \quad (13)$$

$$p_2 = \Pr\{g_i(t) = g_2(t)\} = (1 - p)p_t \quad (14)$$

$$p_3 = \Pr\{g_i(t) = g_3(t)\} = p(1 - p_t) \quad (15)$$

$$p_4 = \Pr\{g_i(t) = g_4(t)\} = (1 - p)(1 - p_t) \quad (16)$$

For a purely random source, the transition density  $p_t$  can easily be shown to be

$$p_t = 2p(1 - p) \quad (17)$$

Inserting the terms from Eqs. (6) through (16) into Eq. (1) and carrying out the necessary mathematics shows the power spectral density  $S_d(f)$  for the asymmetric NRZ data stream generated by a purely random source to be

$$\begin{aligned} S_d(f) = & (A^2 T_s) \frac{\sin^2(\pi f T_s)}{(\pi f T_s)^2} [a_1(p_t) + a_2(p, p_t, \eta)] \\ & + (A^2 T_s) a_3(p_t, \eta) \frac{\sin^2(\pi f T_s \eta)^2}{(\pi f T_s)^2} \\ & + \underbrace{(A^2 T_s) \frac{\sin(\pi f T_s)}{(\pi f T_s)^2} [a_4(p, p_t, \eta) - a_5(p, p_t)]}_{\text{continuous spectrum, } S_c(f)} \\ & + \underbrace{A^2 [2p - (1 - \eta p_t)]^2 \delta(f)}_{\text{dc component, } S_{dc}(f)} \\ & + \underbrace{\frac{2A^2}{\pi^2 p_t^2} \sum_{m=1}^{\infty} \frac{1}{m^2} C(m, p, \eta) \delta(f - m R_s)}_{\text{harmonics, } S_h(f)} \quad (18) \end{aligned}$$

where

$$a_1(p_t) = p_t(1 - p_t) [1 + 2(1 - p_t)] - p_t^3 \quad (19)$$

$$\begin{aligned} a_2(p, p_t, \eta) = & [3p_t^3 + p_t(1 - p_t)[1 + 2(1 - 2p)]] \\ & \times \cos^2(\pi f T_s \eta) \quad (20) \end{aligned}$$

$$\begin{aligned} a_3(p_t, \eta) = & p_t(1 + p_t^2 - p_t) \cos^2(\pi f T_s) \\ & + p_t^3 \cos(2\pi f T_s \eta) \quad (21) \end{aligned}$$

$$\begin{aligned} a_4(p, p_t, \eta) = & p_t(1 - p_t)(1 - 2p) [0.5 \cos(2\pi f T_s \eta) \\ & - p \sin(2\pi f T_s \eta)] \quad (22) \end{aligned}$$

binary NRZ data used in this analysis employs the first model presented in [2] (i.e., +1 NRZ symbols are elongated by  $\Delta T_s/2$ , relative to their nominal value of  $T_s$  seconds, when a negative-going data transition occurs; -1 symbols are shortened by the same amount when a positive-going data transition occurs; and the symbols maintain their nominal  $T_s$  seconds when no transitions occur). This article derives an expression for the power spectral density of an asymmetric NRZ data stream. Using this expression, the data bandwidth can be measured as a function of the data asymmetry for a specified transition density.

Based on the derived formula for the power spectral density, this article determines the levels of undesired spectral components caused by the data asymmetry that falls into the carrier-tracking-loop bandwidth and the amount of telemetry SNR degradation as a function of the amount of data asymmetry. The threshold level of the undesired spectral components causing potential interference to the carrier tracking is characterized by the interference-to-carrier power ratio ( $I/C$ ). A general expression for the  $I/C$  as a function of the amount of data asymmetry is derived.

In 1978, Simon et al. investigated the effects of data asymmetry on system performance [2]. Using this technique, the bit SNR degradations were calculated by evaluating the average bit error rate (BER) in the presence of data asymmetry and then computing the additional bit SNR required because of asymmetry to produce the same value of BER when there is no data asymmetry.

This article develops a simple technique to determine the amount of telemetry-bit SNR performance degradation as a function of the amount of data asymmetry for a specified transition density. The technique presented here uses an approach entirely different from that presented in [2]. The results obtained by this technique are then compared with the results presented in [2].

This article begins with the derivation of the power spectral density for an asymmetric NRZ data stream generated by a purely random source. That is followed by an investigation of the effect of the data asymmetry on the data bandwidth. Next, an analysis of the effect of data asymmetry on the performance of the carrier tracking loop is given along with a set of curves used to determine the critical amount of data asymmetry that produces undesired spectral components, which cause harmful interference to carrier tracking. Finally, this article develops a simple technique to assess the telemetry-bit SNR degradation caused by various values of data asymmetry and then compares the results obtained from this technique with

the results presented in [2], and the test results taken in the Electronic Systems Test Laboratory (ESTL).

## II. Derivation of the Asymmetric NRZ Data Power Spectrum

Consider a random quaternary source that every  $T_s$  seconds transmits a signal from the set  $\{g_i(t); i = 1, 2, 3, 4\}$  with probability  $p_i$ . In the following analysis, the signal is assumed to be generated by a purely random source where the signal transmitted in a given signaling interval is independent of those transmitted in previous signaling intervals. If one defines  $G_i(f)$  as the Fourier transforms of the symbols  $g_i(t)$ , for  $i = 1, 2, 3$ , and 4, respectively, then the power spectral density of the random data stream generated by this source is given by [3]

$$\begin{aligned} S(f) = & R_s^2 |p_1 G_1(0) + p_2 G_2(0) + p_3 G_3(0) \\ & + p_4 G_4(0)|^2 \delta(f) + 2R_s^2 \sum_{m=1}^{\infty} |p_1 G_1(mR_s) + p_2 G_2(mR_s) \\ & + p_3 G_3(mR_s) + p_4 G_4(mR_s)|^2 \delta(f - mR_s) \\ & + R_s \left\{ p_1(1 - p_1) |G_1(f)|^2 + p_2(1 - p_2) |G_2(f)|^2 \right. \\ & + p_3(1 - p_3) |G_3(f)|^2 + p_4(1 - p_4) |G_4(f)|^2 \Big\} \\ & - 2R_s \left\{ p_1 p_2 \Re[G_1(f) G_2^*(f)] + p_1 p_3 \Re[G_1(f) G_3^*(f)] \right. \\ & + p_1 p_4 \Re[G_1(f) G_4^*(f)] + p_2 p_3 \Re[G_2(f) G_3^*(f)] \\ & + p_2 p_4 \Re[G_2(f) G_4^*(f)] + p_3 p_4 \Re[G_3(f) G_4^*(f)] \Big\} \end{aligned} \quad (1)$$

Here,  $R_s = 1/T_s$  is the symbol rate (or signaling frequency), the asterisk denotes a complex conjugate, and  $\Re[\cdot]$  denotes the real part.

It is assumed that the first data-asymmetry model presented in [2] (see Fig. 1) is used in the following analysis. For this particular asymmetric NRZ data stream,

$$g_1(t) = \begin{cases} +A & -(T_s/2) < t \leq (T_s/2)(1 + \Delta) \\ 0 & \text{elsewhere} \end{cases} \quad (2)$$

$$g_2(t) = \begin{cases} -A & -(T_s/2) < t \leq (T_s/2)(1 - \Delta) \\ 0 & \text{elsewhere} \end{cases} \quad (3)$$

# The Impact of NRZ Data Asymmetry on the Performance of a Space Telemetry System

T. M. Nguyen

Telecommunications Systems Section

*The telemetry data asymmetry due to rising and falling voltage transitions can cause undesired spectral components at the output of a spacecraft transmitter. The performance of a space telemetry system can potentially degrade because of these undesired components. This article derives an expression for the power spectral density of an asymmetric nonreturn-to-zero (NRZ) data stream. This formula is then used to investigate how the data bandwidth varies with data asymmetry.*

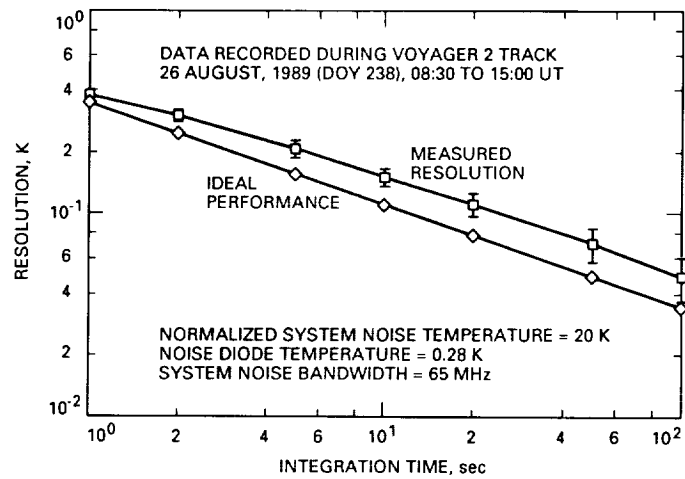
*At the receiver end, the threshold levels of undesired spectral components that fall into the carrier-tracking-loop bandwidth are determined by examining the derived spectral density. Further, based on this formula, a simple technique is developed for the computation of bit signal-to-noise ratio (SNR) degradation due to data asymmetry. The telemetry bit SNR degradations derived using this technique are compared with results obtained previously and with measurements by the Electronic Systems Test Laboratory (ESTL) at the Lyndon B. Johnson Space Center.*

## I. Introduction

The performance of a space telemetry link will be degraded by data asymmetry because of non-ideal rising and falling voltage transitions. The data asymmetry can cause undesired spectral components that fall within the carrier-tracking-loop bandwidth, causing potential interference to the carrier tracking. In addition, these undesirable spectral components can cause telemetry-bit signal-to-noise ratio degradation. Space telemetry links often operate under imposed bandwidth constraints. When two adjacent data channels overlap in the frequency domain, the per-

formance of a space telemetry system can be degraded severely. With these concerns in mind, a communications engineer is interested in determining the power spectral density of the transmitted data.

The Consultative Committee for Space Data Systems (CCSDS) recommends the binary NRZ format for spacecraft-to-ground telemetry data [1]. The transmitted binary NRZ data stream considered in this article is asymmetric in time with a transition density,  $p_t$ , which is less than or equal to  $\frac{1}{2}$ . The data-asymmetry model for the



**Fig. 13. Parkes noise-adding radiometer resolution performance summary.**

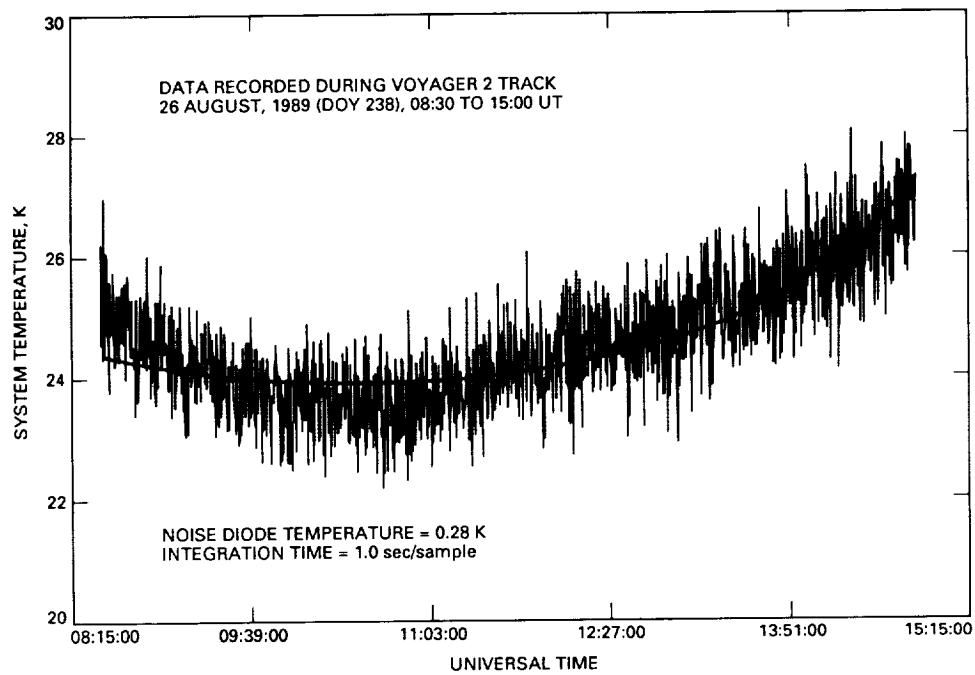


Fig. 11. Logged NAR data and corresponding system noise temperature model.

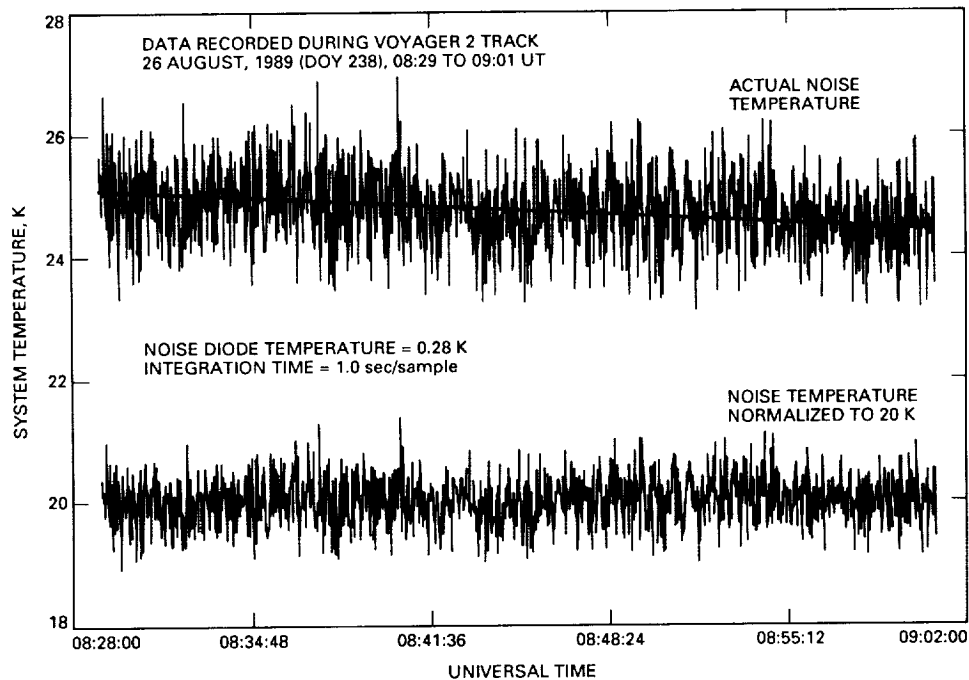


Fig. 12. Transformed A/D data, model, and normalized system noise temperatures.

PFEC>stat d			
PFEC Noise Diode Status (241 12:17:31)			
	RF Ch 1	RF Ch 2	
Noise Diode Temperatures	:	0.223	0.250 Kelvins
	:	0.483	0.500 Kelvins
	:	1.183	1.000 Kelvins
	:	2.089	2.000 Kelvins
	:	3.938	4.000 Kelvins
	:	9.032	8.000 Kelvins
	:	52.382	50.000 Kelvins
	:	61.414	58.000 Kelvins
Noise Diode Calibration	:	Calibrated	Not Cal'd
PFEC>			

Fig. 9. Noise diode status display.

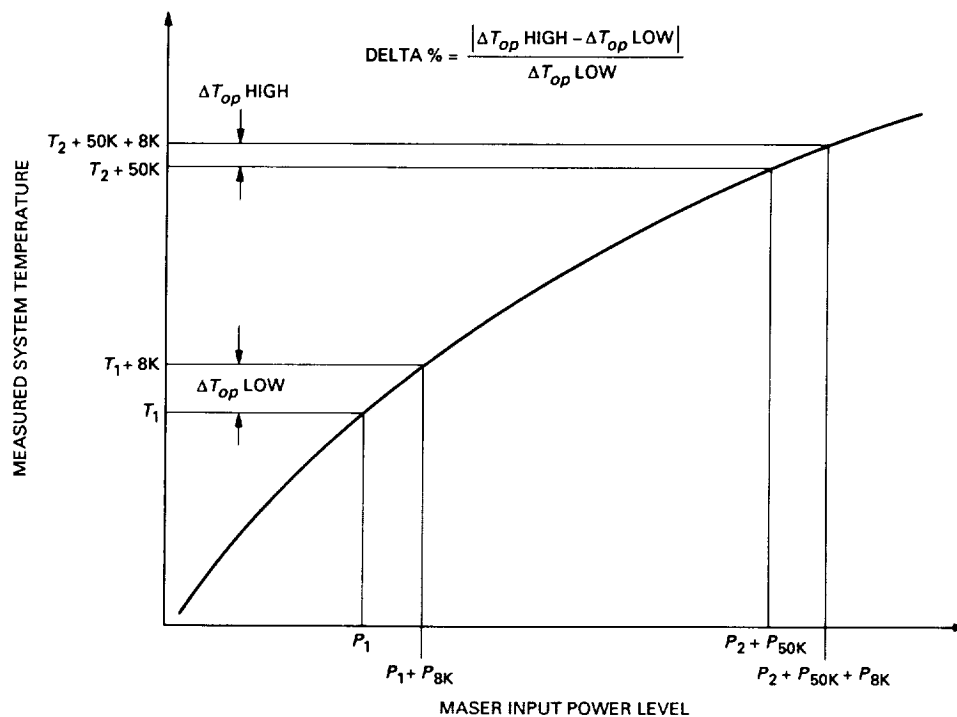


Fig. 10. Parkes radiometer "Delta" command.

PFEC>stat m

PFEC Temperature Measurement Status (236 13:27:54)

	RF Ch 1	RF Ch 2	
System Temperature :	15.328	15.037	Kelvins
Measurement Resolution :	0.025	0.021	Kelvins
Ambient Load Temperature :	297.860	300.163	Kelvins
System Gain Factor :	12.823	8.983	nW/Kelvin
Gain Factor Age :	00:02:38	00:13:25	
Maser Input Noise Temp :	4.05	3.92	Kelvins
System Follow-up Noise Temp :	0.13	0.29	Kelvins
Short-Term Gain Instability :	0.03	0.03	dB
Long-Term Gain Instability :	0.10	0.10	dB/hr
TPR Resolution Uncertainty :	0.013	0.283	Kelvins

PFEC>

Fig. 7. Radiometer temperature measurement status display.

PFEC>stat n

PFEC Radiometer Status (238 11:55:08)

Parkes NAR Command Mode :	Configuration
Radiometer Operating Status :	Running on Channel 2
Radiometer Operating Mode :	Total Power
RF Assembly Detector Mode :	Linear Response
Integration Time Criteria :	Sampling Rate
Measurement Sample Time :	10.0 Hz
Measurement Resolution :	0.030 Kelvins
D/A Maximum Output Temperature :	27.50 Kelvins
D/A Zero-Volt Temperature/Gain :	25.00 Kelvins / 0.500 Kelvins/Volt
D/A Minimum Output Temperature :	22.50 Kelvins
Noise Diode Selection :	51.263 Kelvins, Manually
Noise Diode Modulation :	Automatically Controlled
RF Assembly Internal Attenuator :	-3.832 Volts, Auto-Adjusted

PFEC>

Fig. 8. Radiometer configuration status display.

```

PFEC>help n
PFEC Front End Controller NAR Help...
The format for user input is: COMMAND PARAMETER

  Cmnd  Description                      Parameter
  -----
PNAR    Set Parkes NAR Command Mode.    [OFF],[CFG],[OPR]
NRCV    Set NAR Input Source.            [1],[2],[TRM]
NMOD    Set Total Power or Noise-Adding Mode. [TPR],[NAR]
DMOD    Set Square-Law Detector Operating Mode. [LIN],[DSN] (Debug)
DIOD    Set NAR Diode Temperature.       [A]uto,[OFF],[n] Kelvins (NAR)
DACZ    Set NAR D/A Zero-Volt Temperature. [A]uto,[n] Kelvins
DACG    Set NAR D/A Temperature Gain.     [A]uto,[n] Kelvins/Volt
NOIS    Turn Noise Diodes On or Off.      [A]uto,[OFF],[ON] (Debug)
NATN    Set RF Assy Internal Attenuator.  [A]uto,[n] Volts (Debug)
TMAS    Set Maser Input Noise Temperature. [1],[2] [n] Kelvins
TFOL    Set System Follow-up Noise Temperature. [1],[2],[M]onRcvr [n] Kelvins
GINS    Set Maser Gain Instability.      [S]hort-term [n] dB,
                                           [L]ong-term [n] dB/hr

```

PFEC>

Fig. 5. Radiometer configuration help menu.

```

PFEC>stat c
PFEC Configuration Status (237 07:32:14)          Configuration is UNLOCKED

      53.36K/MOD---+
      |
Horn/RCP-----+-----Maser 1-----+-----X-----NAR
      |                                     |
      8425 MHz/WB/PM---+-----+-----X-----Mon Rcvr
      |                                     |
Ambient Load-----+-----Maser 2-----+-----X-----RF 2
      |                                     |
      ND OFF/OFF---+                                     8100 MHz

```

Local CRT has command locking privileges.

PFEC>

Fig. 6. Front-end configuration status display.

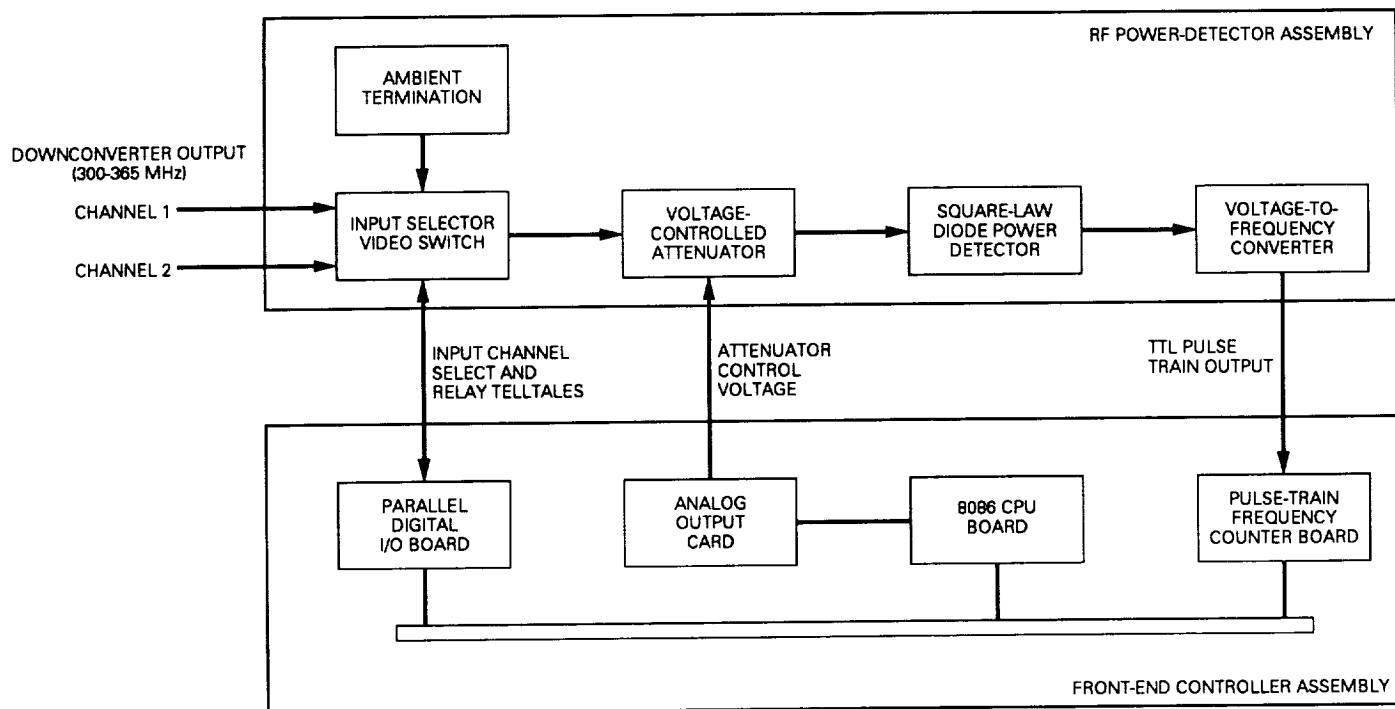


Fig. 3. Parkes radiometer RF power detector.

```

PFEC>help m
PFEC Front End Controller Measurement Help...
The format for user input is: COMMAND PARAMETER

  Cmnd  Description                                     Parameter
  -----
YFAC    Perform Y-factor Measurement.                  [1],[2] [L]eft,[R]ight
GAIN    Compute Maser Gain Profile.                    [1],[2]
PMTR    Operate Power Meter.                           [M]easure,[C]alibrate
CCRH    Chart CCR Performance History.                 [1],[2] [MMddHHmm] [MMddHHmm]
SYST    Measure System Temperature (NAR).              [],[OFF],[C]ont,[D]isp [n] sec
NRES    Set NAR Sample Resolution.                     [n] Kelvins
NRAT    Set NAR Sample Rate.                           [n] Hz
DRAT    Set NAR Diode Switching Rate.                  [n] Hz           (Debug only)
NSAM    Set Number of Resolution Samples.              [n]
CALG    Calibrate System Gain Factor.                  [],[n] nW/Kelvin (CFG only)
CALD    Calibrate Noise Diodes.                       [ALL],[n] Kelvins (CFG only)
DLTA    Measure System Linearity Delta.                [S]ky,[A]mbient,[F]ull Range

PFEC>

```

Fig. 4. Parkes FEC/NAR measurement help menu.

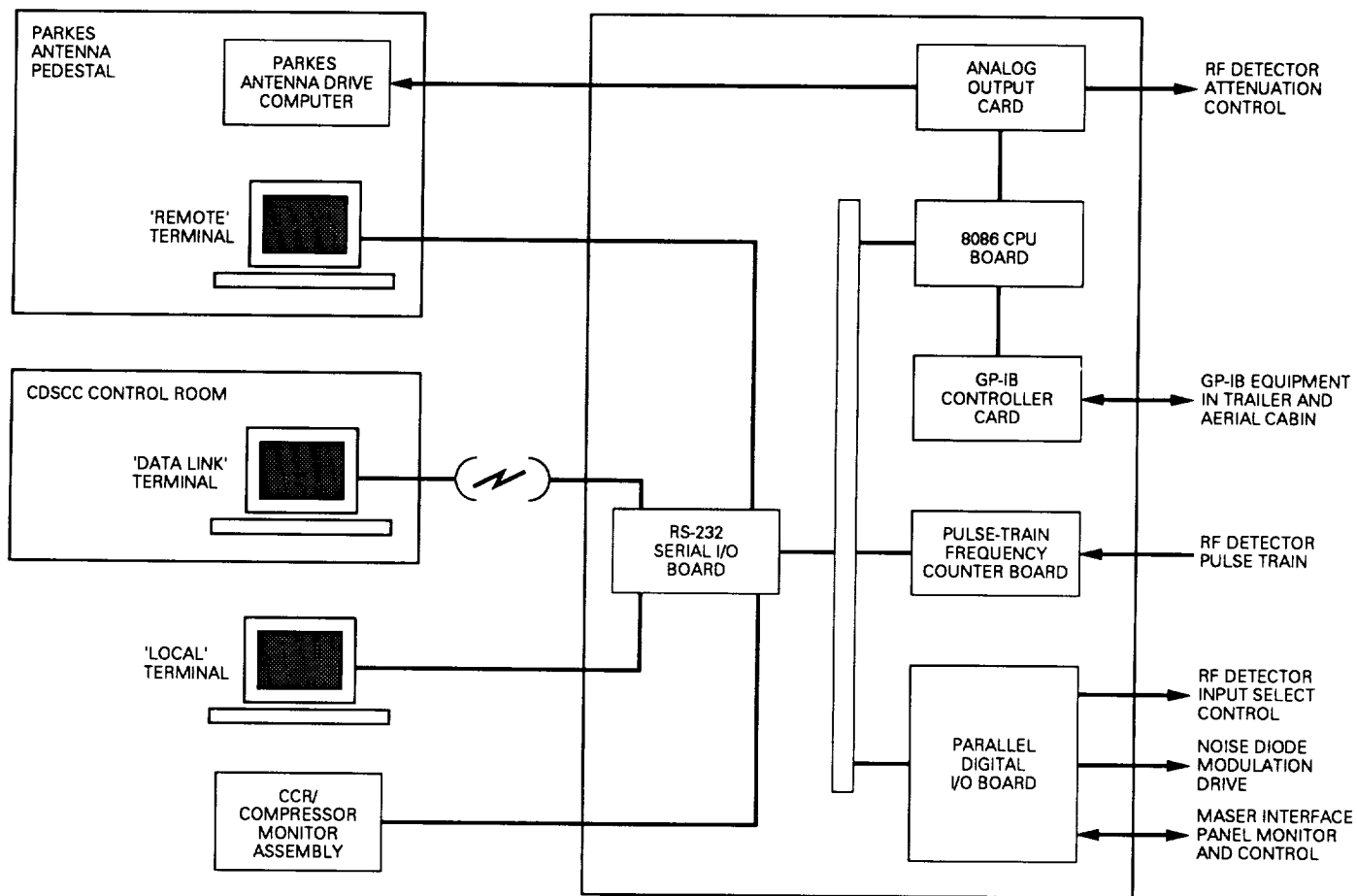


Fig. 2. Parkes front-end controller assembly.

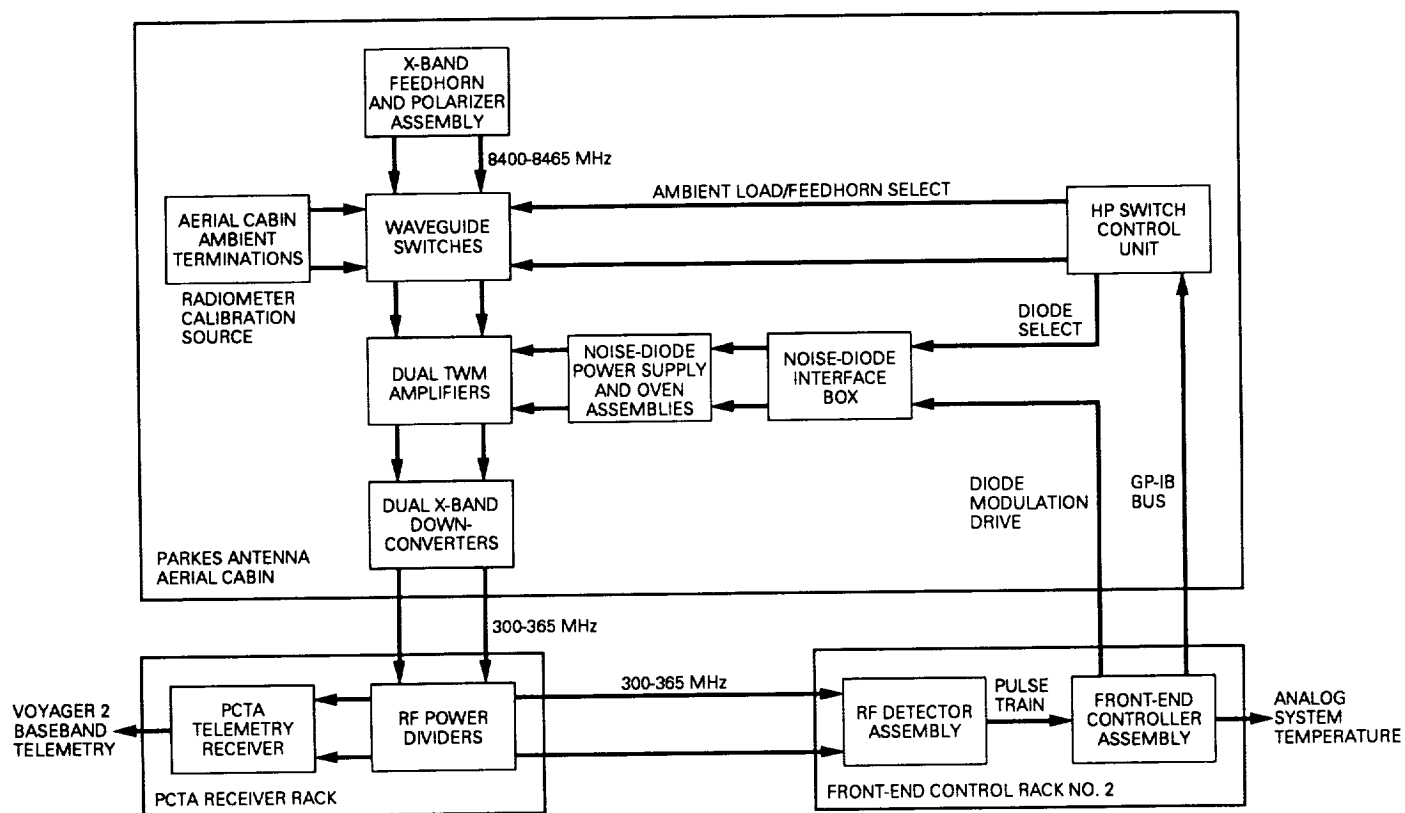


Fig. 1. Parkes front end (radiometer emphasized).

**Table 2. Detail of Parkes noise-adding radiometer resolution performance summary**

Data block	Sample integration time						
	1 sec	2 sec	5 sec	10 sec	20 sec	50 sec	100 sec
1	0.384	0.312	0.215	0.162	0.127	0.088	0.060
2	0.351	0.280	0.188	0.140	0.098	0.058	0.042
3	0.377	0.308	0.211	0.156	0.127	0.088	0.067
4	0.459	0.352	0.268	0.189	0.140	0.097	0.076
5	0.389	0.314	0.223	0.166	0.119	0.076	0.046
6	0.374	0.303	0.207	0.156	0.122	0.079	0.057
7	0.367	0.292	0.192	0.140	0.094	0.058	0.042
8	0.384	0.307	0.216	0.158	0.111	0.077	0.050
9	0.385	0.310	0.208	0.153	0.118	0.058	0.038
10	0.376	0.298	0.212	0.139	0.097	0.063	0.042
11	0.372	0.306	0.200	0.146	0.102	0.066	0.037
12	0.366	0.295	0.201	0.142	0.101	0.063	0.046
Total no. data points	20241	10118	4043	2019	1008	399	196
Average resolution, K	0.382	0.306	0.212	0.154	0.113	0.073	0.050
Standard deviation of resolution	0.02	0.01	0.02	0.01	0.01	0.01	0.01
Ideal performance, K	0.355	0.251	0.159	0.112	0.079	0.050	0.036

**Table 1. Parkes radiometer specifications**

Parameter	Specification
General	
Number of receiver channels supported	Two, switchable
Signal sources (per receive channel)	Antenna feedhorn; aerial cabin ambient termination
Radiometer operating modes	Total power; noise-adding
Nominal system noise temperature range	15 to 400 K
RF detector	
Detector dynamic range	13.5 dB
Nominal input noise power level	-45 dBm
Input frequency bandwidth	65 MHz
Internal attenuator control	Manual or automatic
Internal attenuator range	0 to 50 dB
Detector operating modes	Dynamic first-order fit over subrange of detector response; static second-order fit over entire range of detector response
Noise diodes	
Diode selection control	Manual or automatic
Automatic diode selection criteria	<0.1 dB added noise
Nominal diode temperatures	0.25, 0.5, 1, 2, 4, 8, 50, (50+8) K
Diode modulation control	Manual or automatic
User-specifiable modulation frequency range	0.01 to 100 Hz
Sampling	
Sampling criteria	Integration time; sample resolution
User-specifiable sample integration time range	0.01 to 100 seconds
User-specifiable sample resolution range	0.001 to 10 K/sample
Sample output	
Measurement modes	Manual single-sample; continuous in background
Sample output modes	Display and analog output; analog output only
User-specifiable sample display interval	1 to 15 seconds/sample
Samples used in resolution calculation	None or 2 to 100
Analog output	
Output voltage range	-5 to +5 volts
Output resolution	2.4 mV (12-bits)
User-specifiable zero-volt noise temperature	0 to 500 K
User-specifiable noise temperature/voltage ratio	0.001 to 100 K/V
Calibration	
Total power radiometer mode	System gain factor
Noise-adding radiometer mode	Noise-diode temperatures
Linearity measurement points	Feedhorn/feedhorn+50 K; ambient/ambient+50 K; feedhorn/ambient+50 K

The FEC design maximized the use of existing front-end hardware, implemented in 1985 by ESA for the Uranus encounter, by adding special features and capabilities. In particular, a new noise-adding radiometer design was added that departed from standard DSN design.

Much of the design effort for the NAR focused on performance improvements and operational innovations necessitated by project requirements. The FEC/NAR performed without problems during the encounter and was removed from the Parkes antenna in October 1989.

## Acknowledgments

The author would like to thank C. Stelzried for providing suggestions on NAR functionality and operation. Many thanks also to J. Loreman and D. Trowbridge for their support during development at JPL and to G. Baines and R. Jenkins for their support during on-site installation and testing. L. Hileman contributed a first draft coding of the FEC software.

## References

- [1] D. L. Trowbridge, J. R. Loreman, T. J. Brunzie, and B. Jenkins, "An 8.4-GHz Dual Maser Front End for Parkes Reimplementation," *TDA Progress Report 42-93*, vol. January–March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 214–228, May 15, 1988.
- [2] D. L. Trowbridge, J. R. Loreman, T. J. Brunzie and R. Quinn, "An 8.4-GHz Dual-Maser Front-End System for Parkes Reimplementation," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 301–319, February 15, 1990.
- [3] L. Fowler and M. Britcliffe, "Traveling-Wave Maser Closed Cycle Refrigerator Data Acquisition and Display System," *TDA Progress Report 42-91*, vol. July–September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 304–311, November 15, 1987.
- [4] C. D. Bartok, "Analysis of DSN PPM Support During Voyager 2 Saturn Encounter," *TDA Progress Report 42-68*, vol. October–December 1981, Jet Propulsion Laboratory, Pasadena, California, pp. 139–150, February 15, 1982.
- [5] P. D. Batelaan, R. M. Goldstein, and C. T. Stelzried, "A Noise-Adding Radiometer for Use in the DSN," *JPL Space Programs Summary 37-65*, vol. 2, Jet Propulsion Laboratory, Pasadena, California, pp. 66–69, September 30, 1970.
- [6] T. J. Brunzie, "A Noise-Adding Radiometer for the Parkes Antenna," *TDA Progress Report 42-92*, vol. October–December 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 117–122, February 15, 1988.
- [7] M. S. Reid, R. A. Gardner, and C. T. Stelzried, "The Development of a New Broadband Square Law Detector," *JPL Technical Report 32-1526*, vol. XVI, Jet Propulsion Laboratory, Pasadena, California, pp. 78–86, August 15, 1973.
- [8] C. T. Stelzried, "Non-Linearity in Measurement Systems: Evaluation Method and Application to Microwave Radiometers," *TDA Progress Report 42-91*, vol. July–September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 57–66, November 15, 1987.

each data point and add a constant; however, this method is not sufficient. The problem is that sample resolution is a function of the system temperature, which not only requires that the mean be normalized, but that the variance be normalized as well.

The proper conversion equation can be determined by starting with the equation expressing ideal noise-adding radiometer resolution performance [5,6]:

$$\sigma_T = \frac{2T}{\sqrt{tB}} \left( 1 + \frac{T}{T_D} \right)$$

where

$\sigma_T$  = theoretical best resolution

$T$  = system temperature during measurement, K

$T_D$  = noise-diode temperature, K

$t$  = sample integration time, sec

$B$  = noise bandwidth as seen by RF power detector, Hz

A simplification can be made in this case, where  $T \approx 25$  K and  $T_D = 0.28$  K.

Then  $T/T_D \gg 1$  and the above equation simplifies to

$$\sigma_T = \frac{2T^2}{T_D \sqrt{tB}}$$

For samples taken with a constant integration time, this can also be expressed as

$$\frac{\sigma_T}{T^2} = \frac{2}{T_D \sqrt{tB}} = \text{constant}$$

This relationship provides the means for properly normalizing the mean and variance in the data set, provided an estimate of the true system temperature is known for every sample:

$$T_i'' = \frac{(T_i' - \mu_i)}{\mu_i^2} * T_N^2 + T_N$$

where

$T_i''$  = normalized system temperature sample  $i$

$T_i'$  = transformed A/D converter sample  $i$

$\mu_i$  = mean or "true value" of  $T_i'$  (from model)

$T_N$  = mean normalized system temperature

The mean at each point is estimated by modeling the data using the same method described earlier for converting the A/D data. Figure 12 shows a half-hour segment of A/D data that has been converted to its corresponding system temperature values and plotted, together with its modeled mean and then after normalization to a temperature of 20 K.

A total of 12 half-hour segments of data, each consisting of approximately 1700 one-second integrations, were analyzed for DOY 238. Once each segment had been normalized to a consistent mean and variance, it was reduced to seven resolution figures corresponding to integration times of 1, 2, 5, 10, 20, 50, and 100 seconds. Determining resolution for one-second integration times was straightforward, requiring only the calculation of the standard deviation of each block of points. Resolution for other integration times required additional manipulation of the data.

To obtain data for longer integration times, it was necessary to average together contiguous groups of one-second data points. This technique manually duplicates what the NAR does automatically when integrating for long periods. One consequence of doing so is that the resulting block of data is proportionally smaller than the original, which increases the variance in its calculated resolution.

Each of the half-hour data segments was expanded in this manner six times to produce a total of seven blocks of data. Next, the standard deviation of all seven blocks was calculated for each segment. The resulting twelve values associated with each integration time were then reduced statistically to generate the final performance results. These results, together with the theoretical curve, are tabulated in Table 2 and plotted in Fig. 13.

Although the Parkes NAR performance curve in Fig. 13 does not match the ideal performance curve for a NAR with equivalent characteristics, it should not be expected to do so. The data from which this curve was derived were not well-suited for performance measurements, having been taken during cloudy weather while the antenna was in motion. These conditions will degrade measured resolution, biasing the performance figures. The actual performance of the Parkes NAR should lie somewhere between the measured and ideal curves.

## VI. Summary

The Parkes front-end controller was designed as a cost-effective means of providing a DSN interface for support by the Parkes antenna during the Neptune encounter.

results. The conditions under which the data were taken were, however, significant: poor weather, a moving antenna, and the use of the smallest noise diode, which has the greatest variance in its calibrated value. These factors degraded the measured radiometer resolution, unfavorably biasing the final performance statistics.

### C. Parkes Data Reduction

The first step toward reducing the encounter radiometric data was the transformation of the A/D values back to the original system temperature values. Fortunately, while the NAR was operating, the FEC was instructed to display a system temperature sample every 15 sec on the three control terminals. Because the "data link" terminal at the Canberra complex (an IBM PC) was programmed to log all terminal input/output in a large internal buffer, these system temperature values were captured, recorded on disk, and made available along with the A/D converter values.

This NAR log data set provided the information necessary for reconstructing the original noise temperature values computed by the Parkes NAR. (The data sets could not serve as a basis for NAR performance analysis because 14 out of every 15 samples had been lost; contiguous data were needed.)

Of the three A/D and four NAR log data sets returned to JPL on disk, only day-of-year (DOY) 238—the day following Neptune closest approach—existed electronically in both forms. It is this data set that forms the basis for the performance figures reported in this article.

Because each A/D data set was produced from a series of linear transformations (assuming linear responses from the associated hardware), the transformation back to system temperature values was also a linear process:

$$T'_i = mC_i + b \quad (i = 1, 2, 3, 4, \dots, n)$$

where

$T'_i$  = transformed A/D converter sample  $i$   
(best estimate of the original NAR  $T_i$ )

$C_i$  = A/D converter sample  $i$

$m, b$  = transformation parameters

The transformation parameters were determined by modeling the A/D and NAR log data sets, then substituting the model parameters into the above equation and solving. Under more ideal measurement conditions, the

models would be simply the means of the two data sets. However, because the antenna was in motion during the track, the varying length of the signal path through the atmosphere resulted in a constantly changing baseline system temperature. This required that a more complicated model be used.

In the simpler case, the system-temperature model consists of a constant term. Movement of the antenna introduces a second, time-varying component that accounts for the change in atmospheric signal path length. To simplify calculations, a plane-parallel model of the atmosphere was used; the relative increase in path length compared to zenith could then be expressed more simply as the secant of the zenith angle. Zenith angle can be easily computed for any time of day using the coordinates of the source, the antenna, and standard astronomical tables.

The resulting system temperature models are then

$$C_i = C_C + C_S \sec(\psi_i)$$

$$T_i = T_C + T_S \sec(\psi_i)$$

where

$C_i, T_i$  = model values for each data set

$C_C, T_C$  = model constant terms

$C_S, T_S$  = model scaling factors for time-varying term

$\psi_i$  = antenna zenith angle at time  $i$

Model parameters were determined for each set using the least-squares method with the data values as the dependent variable and the corresponding  $\sec(\psi_i)$ 's as the independent variable.

Figure 11 depicts the NAR log data set for most of DOY 238 and its corresponding model curve. Imperfections in the fit indicate other physical processes (primarily weather) that could not be modeled, but do not significantly affect the transformation of the A/D data set. Once the models were constructed and the transformation parameters generated, the entire A/D data set for DOY 238 was converted to system temperature values. Because a necessary condition for reducing radiometric data to resolution figures is that the set have a constant mean, an additional manipulation was needed.

One method of normalizing the mean of the system temperatures would be to subtract the model value from

Resolving capability depends on several factors: sample integration time, noise-diode temperature and switching rate (for NARs), the bandwidth of the detected signal, and the system temperature itself. As a result, radiometer performance is better expressed as temperature resolution versus integration time for a specified noise-diode temperature at a normalized system temperature.

In order to collect the data needed for radiometer performance analysis, it is necessary to fix some of the variables mentioned above. One of them, the system noise bandwidth, is usually fixed by the hardware. Another involves holding the operating temperature of the system constant during data taking, preferably at as low a temperature as possible. This is because resolution varies inversely with system temperature, and high system temperatures (such as those obtained when viewing the antenna ambient termination) can mask the performance capability of a sensitive radiometer. One solution is to switch the antenna front end to a cold load constructed from a waveguide termination immersed in a liquid nitrogen dewar. (Liquid helium would, of course, be even better.)

For convenience, DSN radiometer performance tests are usually conducted using a clear sky as a source with the antenna held steady, preferably at zenith to minimize the path length through the atmosphere. Under ideal atmospheric conditions, the sky at X-band appears as a constant source radiating at approximately 8.5 K. However, because the atmosphere contributes to the system operating temperature, and it is not constant, the system temperature does not stay constant. The most significant contributing factor at frequencies used by the DSN is water vapor. Unfortunately, even a sky that appears clear may still contain large, variable amounts of water vapor, making this technique only suitable for short test runs at best.

To obtain a thorough picture of radiometer performance under a wide variety of operating conditions, system temperature data need to be collected for several integration times and, in the case of NARs, using a variety of noise diodes. Calculating the standard deviation of each group of measurements yields resolution as a function of integration time and diode. Reducing all the data and plotting yields a family of curves that can be compared to theoretical curves.

Fortunately, the data collection process can be simplified. By operating the radiometer at the shortest integration time of interest and taking a large number of samples, it is possible to integrate groups of samples by hand to obtain data points corresponding to longer integration times. Data sets can be normalized to any desired system tem-

perature, and performance with other noise diodes can be inferred from comparison to theoretical curves.

In interpreting a plot of radiometer performance curves, two things should be examined. The first, of course, is how closely the measured values compare to the corresponding theoretical curve. The second is not so obvious: the location of the theoretical curve itself. This curve is not a universal plot true of all radiometers; it is constructed for the characteristics of that particular radiometer and the conditions under which the data were gathered. It is entirely possible for a radiometer whose performance curve lies well above its theoretical curve to outperform a second system whose performance nearly equals theory.

## **B. Parkes NAR Performance Data**

It is unfortunate that only minimal time was made available for system performance testing after the final version of the Parkes NAR was delivered and installed on the antenna. Immediately after encounter, the radiometer was disassembled and returned to JPL along with the rest of the X-band front-end equipment. As a result, inadequate performance data were obtained.

However, the Parkes NAR was designed to be used for spacecraft tracking as well as antenna-pointing calibration; consequently, the NAR was operated continuously while the station tracked Voyager 2 throughout encounter week to provide the Voyager radio science investigators with baseline system noise temperature data. Although this type of data is not ideally suited for radiometer performance measurements, a method was found to derive from it a single performance curve for the NAR. The development of this method and the final results are described next.

During the Voyager encounter tracks, the Parkes NAR was instructed to integrate samples for a period of one second using the 0.25-K noise diode. The antenna drive computer, already configured to digitize the NAR analog output to support antenna pointing calibration, was reprogrammed to receive the system temperature samples and record them to floppy disk throughout each track.

These data sets were recorded by the Parkes computer as raw analog-to-digital (A/D) converter values, not as operating system noise temperatures. Unfortunately, the proper conversion factors were not available to reconstruct the original system temperatures and the continuity of the data was disrupted by gaps due to tracking update activities. These factors necessitated additional steps in the data analysis, but did not impact the accuracy of the final

tional numeric parameter to the command allowed manual setting of the temperature of the active noise diode.

An estimate of system linearity and NAR health could be obtained with a command measuring the additive noise temperature of a medium-power diode at two different system noise temperature levels and comparing the results. Typically, the radiometer first measured the baseline-feedhorn system temperature in the TPR mode and then the temperature with the medium diode turned on. Secondly, the TPR measured system temperature using the ambient load with the highest noise diode on and then with both the highest diode and the medium diode on. The differences between the two measurements at each operating point yielded a derived medium-diode temperature. The percent difference between the two derived values was reported to the operator as a measure of system linearity (Fig. 10).

Measurement resolution for a NAR is a function of integration time, diode temperature, system temperature, and noise bandwidth. For systems that are not gain-stable, NAR resolution is also affected by diode modulation rate. NARs are gain-insensitive only when individual measurement periods are short enough that the system gain does not change appreciably during sample integration. For short integrations, a pair of relative measurements eliminates the effects of a changing baseline power level. Slow diode modulation rates allow gain instability effects to appear and degrade resolution. High diode switching rates cause the controller to spend most of its time on processing overhead and very little time collecting data. In either case, measurement resolution decreases. A diode switching rate command provided the ability to adjust the diode modulation rate to obtain the best results (typically 10–15 Hz for Parkes and other DSN antenna systems).

In some cases, it was desirable to manually modulate the noise diodes or set the RF power-detector attenuator level. A manual command allowed the operators to force the diodes to turn on or off, or to return automatic control to the FEC. Another command allowed the operator to override the automatic adjustment of the internal attenuator and manually set an attenuation level.

An additional operating mode not normally accessed by the operator affected operation of the RF power detector. One of two methods of determining noise-power level with the detector could be selected. The default mode, referred to as the “linear” mode, used a second-order transfer function (based on original measurements of the equipment) to create a linearized relationship over the entire detector operating range. This relationship was between the RF

power entering the assembly and the power reading computed by the driver routines. As a result, the location of the operating point was not critical; the internal attenuator was always automatically adjusted to center the range of measurable temperatures within the dynamic range of the detector, and then held constant until a change in system configuration was detected. This type of operation was critical for TPR mode use because of the large temperature difference between the ambient load calibration source and the feedhorn source. Without it, the internal attenuator would have to be adjusted to prevent clipping or to maintain a constant operating point, either of which would change calibration baselines and decrease accuracy.

The alternate RF detector operating mode, known as the “DSN” mode, was simpler and patterned after the method used by the DSN PPM NAR. This mode could only be used in conjunction with NAR operation, as it allowed the detector’s internal attenuator setting to be altered, erasing any calibration baseline. The DSN mode dynamically fit a first-order equation to the detector frequency response, thereby restricting power measurements to a small, nearly linear region of the response curve. This was done by adjusting the internal attenuator to maintain a fixed output frequency for the diode off state prior to each measurement. Noise power was computed by dynamically fitting a first-order equation to the off measurement frequency and the detector’s internal ambient termination rest frequency.

## IV. Monitor and Control

Because the Parkes FEC was never required to be used at a Deep Space Communications Complex (DSCC), its command format and monitor displays were not required to follow DSN CMC/LMC interface agreements and convention. Originally, the FEC was to be connected to the PCTA; thus, to maintain compatibility, the FEC command structure and displays were designed to meet PCTA conventions. When the decision was made not to implement the PCTA controller link, the FEC became a stand-alone controller and was connected to the Canberra DSCC via a modem and data link.

## V. Performance

### A. Discussion

Probably the most important performance statistic for judging the quality of a radiometer is the degree to which it can resolve changes in system operating noise temperature. Reporting radiometer performance is not, however, as simple as quoting a measured value of temperature resolution.

the FEC. The name "square-law" refers to the square-law diode used by the detector to measure input signal power: the voltage developed across the diode is proportional (ideally) to the total power in the signal. This voltage is translated into a TTL-level pulse train whose frequency represents signal power; the pulse train forms the output of the assembly. For the 1989 Voyager encounter, the pulse-train frequency was measured using a JPL-designed digital frequency-counter board.

Although the input power-output frequency relationship of the RF detector assemblies was intended to be linear, in practice it more closely resembled a parabola than a straight line. The Parkes NAR and the PPM employ different software techniques to counteract this characteristic, both of which are discussed in the next section. Additional methods of compensating for this and other front-end nonlinear behaviors are described elsewhere [8].

The RF detector assemblies (Fig. 3) provide two interfaces for controlling power measurement: an input selector switch and an RF attenuator. The input selector switch is a parallel TTL-controlled video switch used to select one of up to eight signal channels for measurement. The voltage-controlled attenuator is needed to adjust the signal level at the diode and set the desired baseline output frequency. Attenuation is minimized at  $-7$  volts and increases exponentially as the control voltage approaches a limit of  $+1$  volt.

The voltage-to-frequency units in the detectors were designed to output a frequency of  $100$  Hz with the detector switched to an internal ambient termination. The maximum output frequency that could be obtained with the Parkes unit was  $13.7$  kHz. As a result, the detector provided a dynamic range just exceeding the approximately  $12.5$  dB in input power experienced by the Parkes radiometer during operation. (System temperatures ranged from a zenith clear-weather value of  $20$  K to the antenna-ambient-termination-plus-noise-diodes value of  $360$  K.) This characteristic allowed the detector to be operated without changes in the internal attenuator setting, once the proper operating point was determined and the attenuator set accordingly. As discussed in the following section, this characteristic was crucial for maintaining a consistent calibration baseline.

**3. Operation and Calibration.** Although the Parkes radiometer is referred to as a noise-adding radiometer, the term is somewhat misleading: it could be operated as either a noise-adding radiometer (NAR) or as a total-power radiometer (TPR). In the TPR mode, the system temperature was derived from measurements of to-

tal RF power in the downconverter band. Therefore, system temperature was measured nonintrusively, but was subject to system gain instabilities and required periodic recalibration on the ambient load to maintain accuracy. In the NAR mode, system temperature was derived by Y-factoring power measurements made with the noise diodes modulating on and off. This mode reduced sensitivity to gain changes and provided better resolution when using larger noise diodes, but telemetry was degraded due to the noise injected into the system.

In the NAR mode, the operator could specify which noise diode to modulate by entering its temperature as a parameter. The FEC did not require an exact value; it would find the closest available diode and inform the operator of its choice. The FEC could also automatically monitor the system temperature and select a diode that would not increase the system noise temperature by more than two percent.

In the TPR mode, noise-power measurements were converted into system temperature using the fundamental relationship  $P = GkTB$ . The  $GkB$  value, referred to in FEC displays as the "system gain factor," varied with system gain and had to be calibrated prior to using the TPR, and periodically thereafter. The FEC performed a system gain factor calibration operation only for the maser channel selected. The automated sequence of events involved selecting the antenna ambient termination as the signal source for the receive chain, performing a power measurement with the RF detector, restoring the original configuration, and then solving for the system gain factor. Further system temperature calculations used this value until it was recalibrated at the request of the operator. Adding an optional numeric parameter to the command allowed manual setting of the system gain factor value.

Use of the NAR mode also required a calibration operation. Noise diode reference temperatures can only be measured in situ, and can vary in value as their environment changes, so they generally are measured prior to each use. The process of calibrating the entire noise diode set for the selected maser channel was automated using a calibration-transfer technique. This was done by selecting the antenna ambient load as the signal source, taking NAR Y-factor data using the largest noise diode, computing diode temperature from the data, selecting the feedhorn as the signal source, and measuring the sky temperature using the newly calibrated large diode. Once the sky temperature had been measured, each of the remaining diodes was selected for calibration using the sky temperature as a reference. Once all the diodes were calibrated, the original receive-chain signal source was restored. Adding an op-

via coaxial cable, and used to drive a strip-chart recorder and analog front end on the Parkes antenna computer for pointing calibration and radio science data collection.

**3. Commands.** Commands for the FEC were grouped into menus according to function. Each menu item included the command name, a brief functional description, and a syntactic representation of parameter types. Syntax and range checking was enforced; individual error messages explained how to correct command errors. Each group had a menu and one or more corresponding status displays. Most commands also responded to a "query," which produced a one-line status display for the associated function. FEC menus and status displays that pertained to the radiometer are shown in Figs. 4 through 9.

**4. Interfaces.** The FEC provided cabling interfaces with the front-end equipment via IEEE 488, parallel-transistor-transistor logic (TTL) control lines, RS-232 serial communication lines, and 50-ohm coax. The equipment interfaces are described in [1].

## B. Noise-Adding Radiometer

It is difficult to describe the NAR on a component level; it can more readily be conceptualized as a function of the FEC than as a set of components. Dedicated NAR components included the noise diode assemblies (located on the maser package in the aerial cabin), RF power-detector drawer (located below the FEC in front-end control cabinet no. 2), and its associated frequency-counter board (located within the FEC). These components are described below, along with NAR operation and calibration. A description of the NAR design before it was implemented is detailed in [6]; however, several changes have been made to this design since the report was published. Block diagrams of the NAR are shown in Figs. 2 and 3.

**1. Noise-Diode Assemblies.** Two noise-diode assemblies were recovered from the ESA Parkes front-end system, implemented during the Voyager Uranus encounter in 1986. Each assembly consisted of an oven with three diodes and a power supply. Each diode was capable of three noise levels, depending on the value of its supply current. The diode's power supply assembly contained three independent current sources, one per diode. Any of the three current levels could be selected through three relay inputs and monitored with three relay telltales. A separate TTL-level input for each diode provided instant switching of the supply current, allowing rapid diode modulation.

The relay inputs, telltales, and modulation-drive signals were designed to be monitored and controlled by a PPM

through a PPM NAR controller assembly. But, in this case, the PPM controller was functionally replaced by the FEC, which controlled the noise diodes with an HP 3488A switch/control unit (SCU). This unit was a rack-mounted programmable device that used removable interface cards to monitor and control various functions. Communication with the SCU was provided by an IEEE-488 GP-IB standard bus that connected the FEC with equipment in the trailer and the aerial cabin.

Two SCU bidirectional digital I/O interface cards were used by the NAR to drive the noise diode power supply relays and to sense the relay tell-tales. Each card provided sixteen channels programmed for eight bits input and eight bits output. Although there are potentially nine inputs and nine outputs for each noise diode assembly, only seven diode temperatures were standard on the units: 0.25, 0.5, 1, 2, 4, 8, and 50 K. Each card provided independent closed-loop diode selection for one of the two maser channels.

Because diode modulation requires precisely coordinated switching that the SCU cannot provide when using a remote, shared GP-IB, the SCU was not used for this purpose. A single modulation signal was transmitted separately from the trailer on a coaxial cable and used to modulate all six diodes simultaneously. And because the project requirements for the NAR were limited to pretrack calibration support, no attempts were made to provide the multiple-diode selection and independent modulation schemes sometimes desired for radio science applications.

Noise output was injected through a waveguide coupler between the feedhorn/ambient load switches and the maser input flanges. The maser input flange served as the reference point for calculating system temperatures. The actual power measurements were made at the PCTA telemetry receiver input in the equipment trailer.

**2. RF Power-Detector Assembly.** The Parkes NAR RF power-detector assembly was a borrowed PPM 8.4-GHz (X-band) prototype square-law detector [7]. Originally, as described in [6], plans were made to use digital signal-processing techniques to develop an alternative to the PPM detector (the digital power meter). This choice was motivated by the poor linearity of the PPM device, which results in inaccurate measurements. Because of budgetary constraints, however, the RF power detector was incorporated into the Parkes radiometer design instead.

In typical DSN use, the RF detector is part of a PPM operated by a PPM NAR controller [4]. However, for the Parkes implementation, the detector was operated by

to minimize conflicts in requirements through efficient software design, but some trade-off in performance was necessary. Project requirements that limited the design included the need to use existing PPM components to save time and reduce cost and the need to time-share hardware and software with other front-end functions.

### III. FEC/NAR System Description

The Parkes tracking station (designated DSS 49) was configured for the Voyager Neptune encounter as a two-channel X-band receive-only system, using two cryogenic maser amplifiers. The front-end package, which contained both masers and the microwave feed, was positioned at the prime feed focus in the antenna's aerial cabin. The NAR noise-diode assemblies and support equipment were part of the front-end package. The Voyager telemetry receiver and associated signal-processing equipment were located in a double-wide trailer just outside the antenna pedestal. The associated FEC equipment was divided between these two locations (see Fig. 1) [2].

#### A. Front-End Controller

1. **Description.** The FEC consisted of a multibus computer providing a card cage, power supply, and enclosure with slides. It monitored and controlled the microwave electronics in the antenna front end and provided for two modes of radiometer operation: total-power and noise-adding. A block diagram of the FEC is shown in Fig. 2.

The FEC central processing unit (CPU) was a Multibus-based Intel 8086 single-board computer containing GP-IB and digital-to-analog (D/A) converter piggy-back cards. FEC support boards included two ROM boards, a RAM board, a four-port serial communications board, a 3- by 24-bit parallel communications board, and a JPL-built pulse-train frequency counter board. Most of the hardware interfaces for NAR operation used ports on input/output (I/O) boards shared with other FEC functions. The only additional hardware added to the FEC for NAR operation was the D/A converter card and the pulse-train frequency counter board. Monitor and control signals for the NAR shared the same communications network used by other FEC functions.

The FEC D/A converter provided two analog signals for NAR operation, the control signal for the RF power-detector drawer (RF detector) internal attenuator and the analog system temperature output. A channel on the parallel I/O board provided 16 bits for the RF detector input channel select and confirmation telltales as well as the

diode modulation drive output. Noise diode selection and confirmation was performed via an IEEE-488 GP-IB interface to an HP 3488A switch/control unit located in the aerial cabin.

The pulse-train frequency counter board, a standard PPM design, provided an integrate-and-dump means of measuring the output frequency of the RF detector drawer (which is a function of input power). Once instructed to begin a measurement, the board operated unattended, interrupting the CPU when data collection was completed. By counting input pulses and system clock ticks simultaneously, the board allowed the FEC to collect both timing and RF power data. Dividing the number of input pulses by measurement duration yielded average frequency, which was then further processed to yield input power and system temperature, as will be discussed further in Section B.

2. **Operation.** The FEC was controlled through three CRT terminals, each able to issue commands and receive displays. The "local" terminal was located in one of the front-end control racks in the trailer and was used for front-end maintenance and testing. It also was used exclusively for operation of the NAR during spacecraft tracking. The "remote" terminal was located in the antenna control room on the third floor of the antenna pedestal and was used by Parkes personnel to operate the NAR in support of antenna calibration and X-band radio astronomy observations. The "data link" terminal was located in the control room at the Canberra Deep Space Communications Center (CDSCC) and was used to monitor front-end health and status via modem during pre-encounter periods when the antenna was not being operated by station personnel.

All three of these terminals were able to send commands and monitor front-end status at any time, but only one terminal at a time functioned as the master terminal, possessing the ability to lock the front-end configuration for security during actual tracking operations. The master terminal was selected via a switch on the front of the FEC chassis. All data, commands, and information graphics could be printed on a local printer.

Radiometer data were made available by CRT displays, printouts, and via a time-varying analog signal. The display information (which could be suppressed to prevent interference with other FEC commands) was copied to all three terminals for simultaneous display. Printouts of these displays were made by connecting a printer to the CRT. The analog output, available at the FEC back panel, represented system temperature values as voltage levels. This signal was updated with each temperature sample computed, transmitted to the pedestal third floor

- (2) Measure system operating noise temperatures using the Y-factor technique through the automated waveguide switches, quartz thermometer, dual traveling-wave maser (TWM) control assembly, and power meter.
- (3) Measure gain versus frequency of the TWMs, with storage, retrieval, and display.
- (4) Perform remote-controlled, noise-adding radiometer functions through automation of the noise diodes and the RF detector power meter during antenna calibration. Functions included returning a time-varying analog signal, representing operating noise temperature, and periodic calibration of the noise diodes.
- (5) Chart performance history, provide status updates, and generate alarm messages for the closed-cycle refrigerator/compressor monitor system [3].
- (6) Monitor alarms for the downconverters, monitor receiver, and test signal upconverter, including power supplies, local oscillator power levels, timing, and phase-lock conditions.

The overall functions and components of the Parkes FEC are described in [1], which details the front-end system design before it was implemented. This article will focus primarily on the noise-adding radiometer (NAR) function of the FEC, since adding this capability comprised a major design effort during FEC development.

Typically, NARs at DSN tracking stations are implemented in the form of a precision power monitor (PPM) assembly [4]. But in the case of the Parkes antenna, a PPM was not installed for the Neptune encounter because of budgetary constraints. Instead, a new NAR design was developed as part of the new Parkes front-end controller.

The NAR met all requirements (see Table 1) and functioned without difficulty during the Voyager Neptune encounter. The FEC/NAR equipment was removed from the Parkes antenna on October 2, 1989 and returned to JPL.

## II. NAR Requirements

As was mentioned, one of the functions performed by the Parkes FEC was radiometric measurement of the operating noise temperature of the receiving systems. System noise temperature data are needed for RF feed focusing, collection of pointing data, and measurement of baseline atmospheric effects. All radiometers determine noise temperature by comparing receiver noise level against a calibrated noise source. The radiometer most commonly used

in the DSN is referred to as a noise-adding radiometer because it periodically adds calibrated noise to the signal during the measurement process [5].

Requirements for the Parkes NAR originated from the need to calibrate the pointing and focus of the antenna after the new front-end package was installed. This type of calibration is performed by pointing the antenna at extragalactic radio sources, quasi-stellar objects that radiate at radio frequencies. The resulting increase in receiver noise level caused by these radio sources appears as a rise in noise temperature that is detected and measured by the radiometer. Optimum antenna focus and pointing calibration is achieved by varying equipment positions and other parameters in the aerial cabin, RF package, and feedhorn until a peak in system noise temperature is obtained.

During pointing calibration, the telescope makes two orthogonal sweeps across the catalog coordinates of a star, digitizing and recording the system temperature as a function of position. The antenna system computer derives a Gaussian curve from each sweep, then mathematically constructs a model that fits these Gaussian curves to previously cataloged information on the flux and position of the star. The difference between the cataloged values and the center of the sweep is the pointing error. The procedure is repeated for several combinations of azimuth and elevation angles to model and systematically eliminate these pointing errors. However, these pointing curves are distorted somewhat by poor gain stability in the receiving system and also by insufficient NAR sample resolution. The stability requirement for the Parkes NAR was 0.1 K during the sampling time (corresponding to a gain stability of 0.02 dB) and the resolution requirement was 0.01 K for returned samples.

Ideally, the computers that control the pointing and focusing of the Parkes antenna would receive a continuous, instantaneous feedback of system temperature with infinite resolution. In practice, each measurement is limited in resolution and requires some minimum amount of integration time before a suitable system temperature sample can be produced. The Parkes NAR was required to return samples at a rate of 10 per second or faster, with 20 samples per second desired. These temperature samples were to be supplied as time-varying analog voltages sent to the control room through coaxial cable.

Other constraints limited design options for the Parkes NAR. Since radiometer sample resolution, sampling rate, system temperature, noise-diode selection, and system hardware characteristics are not independent parameters, they could not be designed as such. Attempts were made

August 15, 1990

N 91 - 11983

# The Parkes Front-End Controller and Noise-Adding Radiometer

T. J. Brunzie

Radio Frequency and Microwave Subsystems Section

*A new front-end controller (FEC) was installed on the 64-m antenna in Parkes, Australia, to support the 1989 Voyager 2 Neptune encounter. The FEC was added to automate operation of the front-end microwave hardware as part of the Deep Space Network's Parkes-Canberra Telemetry Array. Much of the front-end hardware was refurbished and reimplemented from a front-end system installed in 1985 by the European Space Agency for the Uranus encounter; however, the FEC and its associated noise-adding radiometer (NAR) were new JPL designs. Project requirements and other factors led to the development of capabilities not found in standard DSN controllers and radiometers. The Parkes FEC/NAR performed satisfactorily throughout the Neptune encounter and was removed in October 1989.*

## I. Introduction

The 64-meter Parkes Antenna of the Australian Telescope National Facility (or the Parkes antenna) was temporarily used to enhance X-band (8.42 GHz) receiving capability from the Voyager spacecraft during the Neptune encounter in 1989. The Parkes antenna formed part of the DSN Parkes-Canberra Telemetry Array (PCTA) and was outfitted with DSN-compatible hardware, including a microwave front end, telemetry receiver, data recorders, and radio science equipment. Much of the hardware was refurbished and reimplemented from the front-end system developed by the European Space Agency (ESA) for the Uranus encounter in 1986; however, the previous front-end monitor and control system was replaced with a new JPL design. (For an overall description of

the design and installation of the front-end system, see [1,2].) A block diagram of the front-end system is shown in Fig. 1.

The Parkes front-end controller (FEC) centralized and automated the monitoring and control of the front-end system, tying together equipment located in a NASA trailer next to the antenna with hardware located in the antenna's aerial cabin. A block diagram of the FEC is shown in Fig. 2.

The primary tasks performed by the FEC were to

- (1) Monitor and control the antenna waveguide switches and the polarizer drive assembly using a remote switch/control unit and telltale monitoring.

ORIGINAL FILED  
BLACK AND WHITE PHOTOGRAPH

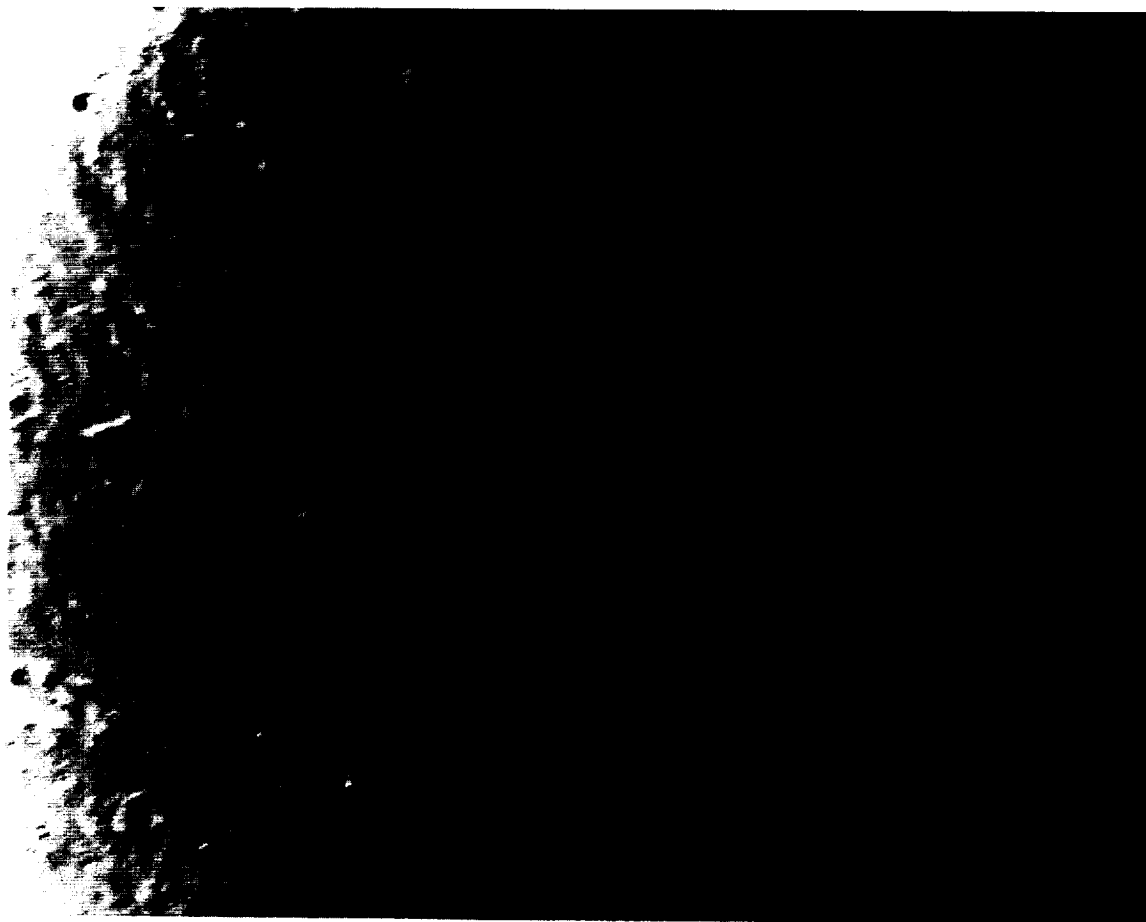


Fig. 9. A raw playback image of Triton received by the VGTA on day-of-year 239 (1989) at 03:59 UT. Data rate = 21.6 kbps; SSNR = 4.2 dB; BER =  $5.4 \times 10^{-5}$ .

EXHIBIT 100  
BLACK AND WHITE PHOTOGRAPH



Fig. 8. The Triton Image shown in Fig. 7 after de-spike processing.

ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH

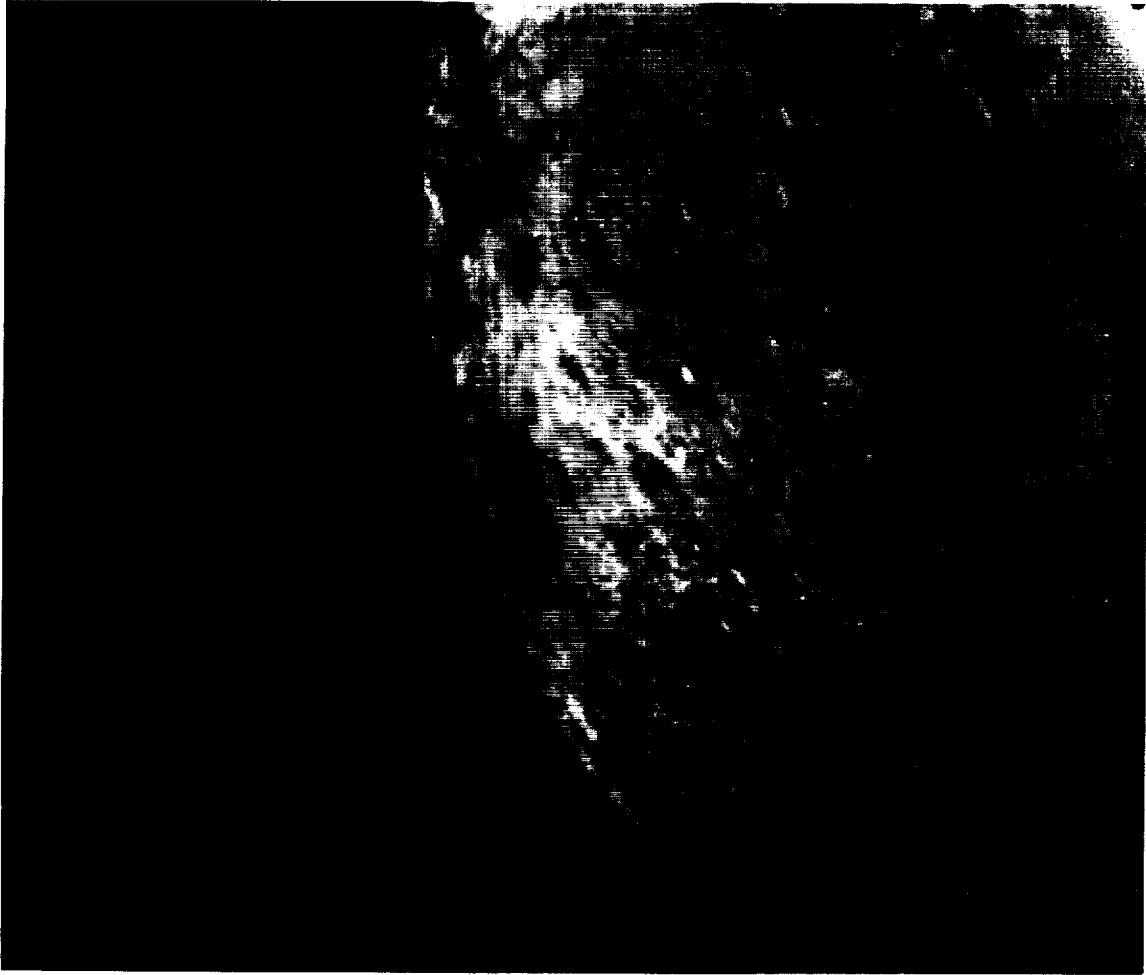


Fig. 7. A raw playback image of Triton, received by the VGTA on day-of-year 239 (1989) at 01:37 UT. Data rate = 21.6 kbps; SSNR = 3.4 dB; BER =  $1.7 \times 10^{-4}$ .

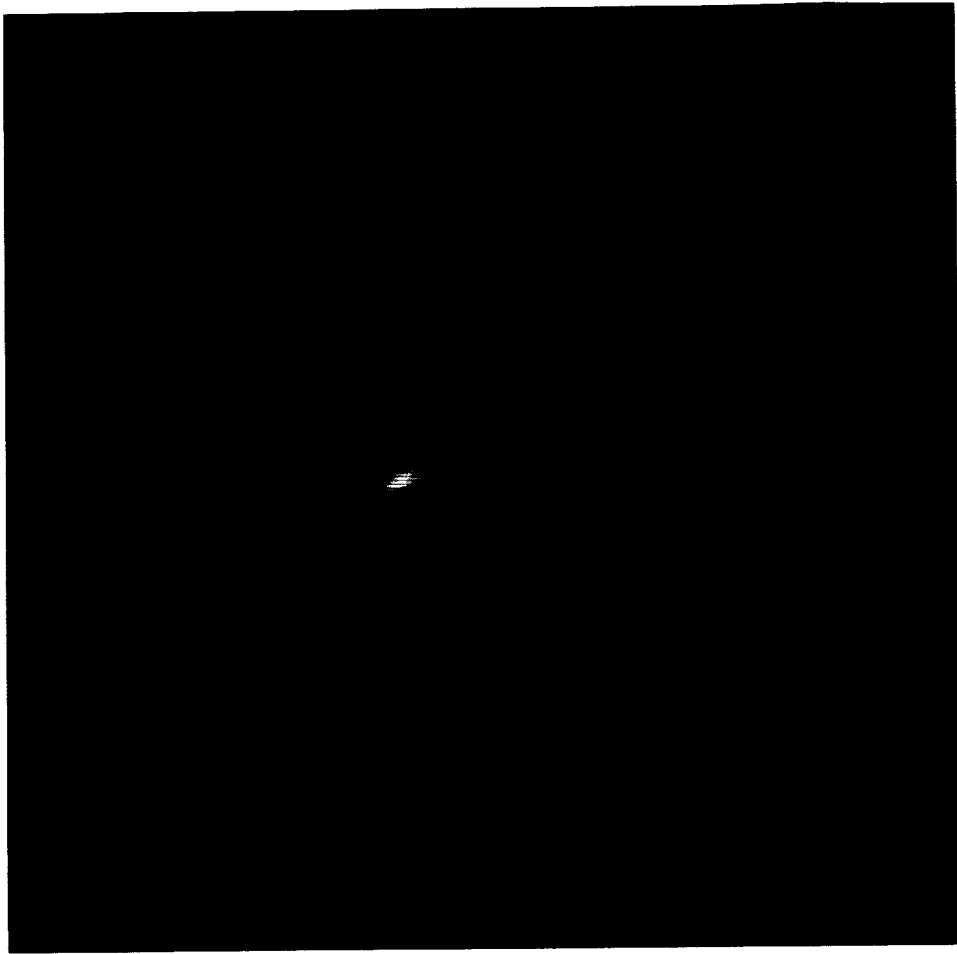


Fig. 6. A Neptune image received by the VGTA on day-of-year 236 (1989) at 04:06 UT. Reed-Solomon coding; data rate = 21.6 kbps; SSNR = 4.4 dB; BER =  $3.7 \times 10^{-5}$ .

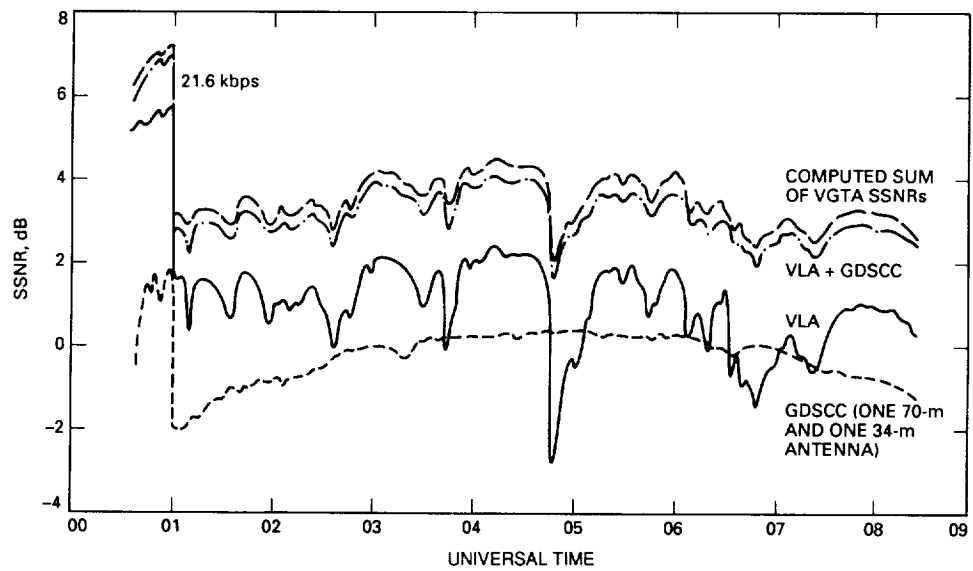


Fig. 4. Constituent symbol SNRs for the VGTA on day-of-year 228 (1989).

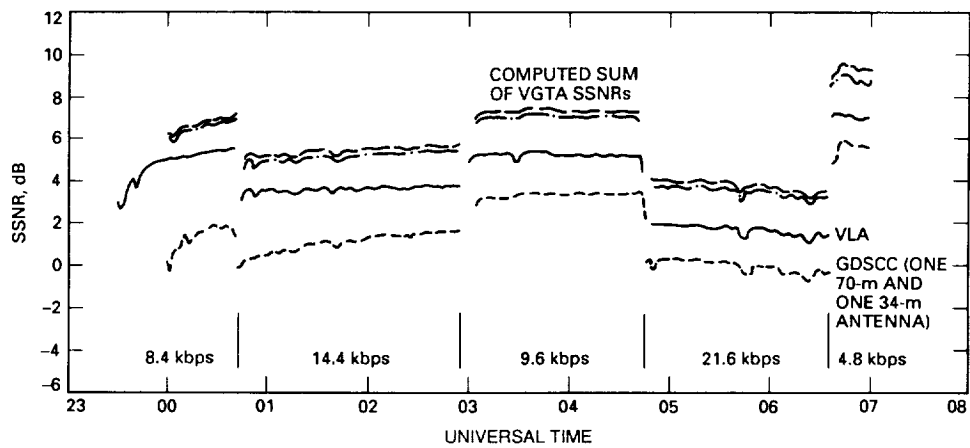


Fig. 5. Constituent symbol SNRs for the VGTA on day-of-year 237 (1989).

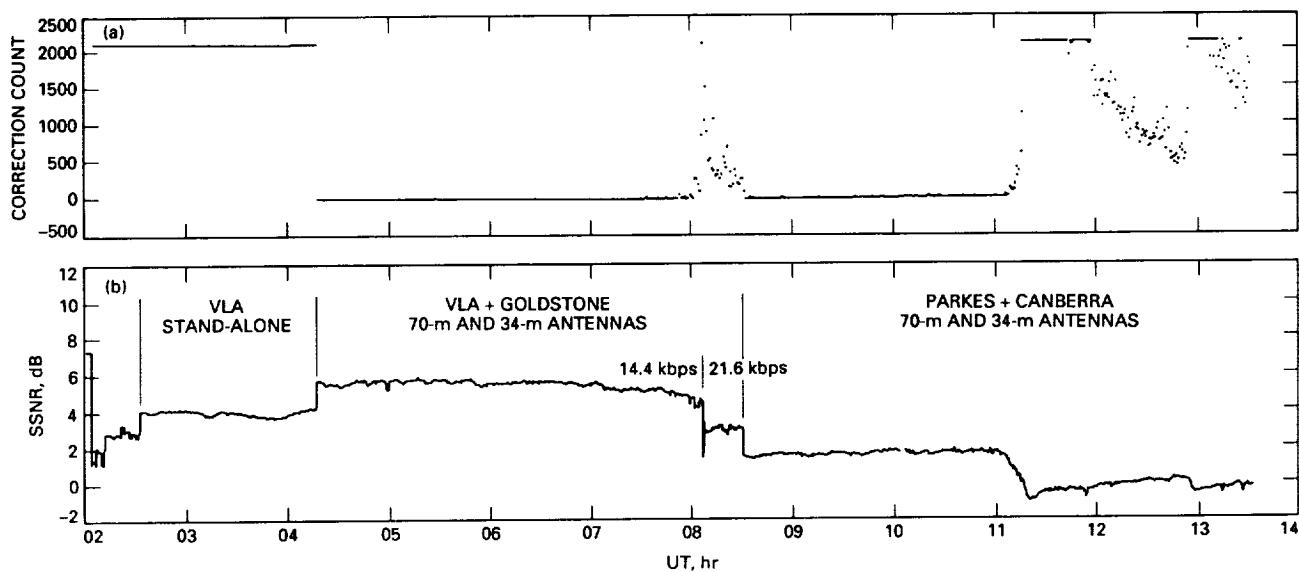


Fig. 2. Voyager data reception by VGTA and PCTA, day-of-year 222 (1989): (a) Reed-Solomon error-correction count, and (b) symbol SNR.

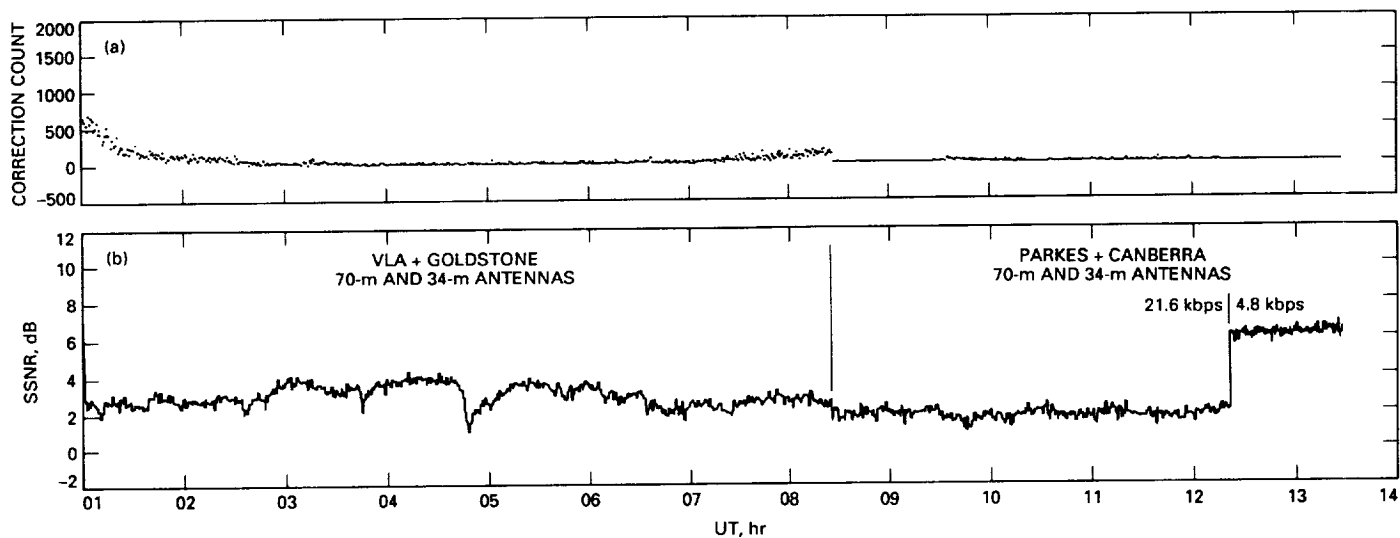


Fig. 3. Voyager data reception by VGTA and PCTA, day-of-year 228 (1989): (a) Reed-Solomon error-correction count, and (b) symbol SNR.

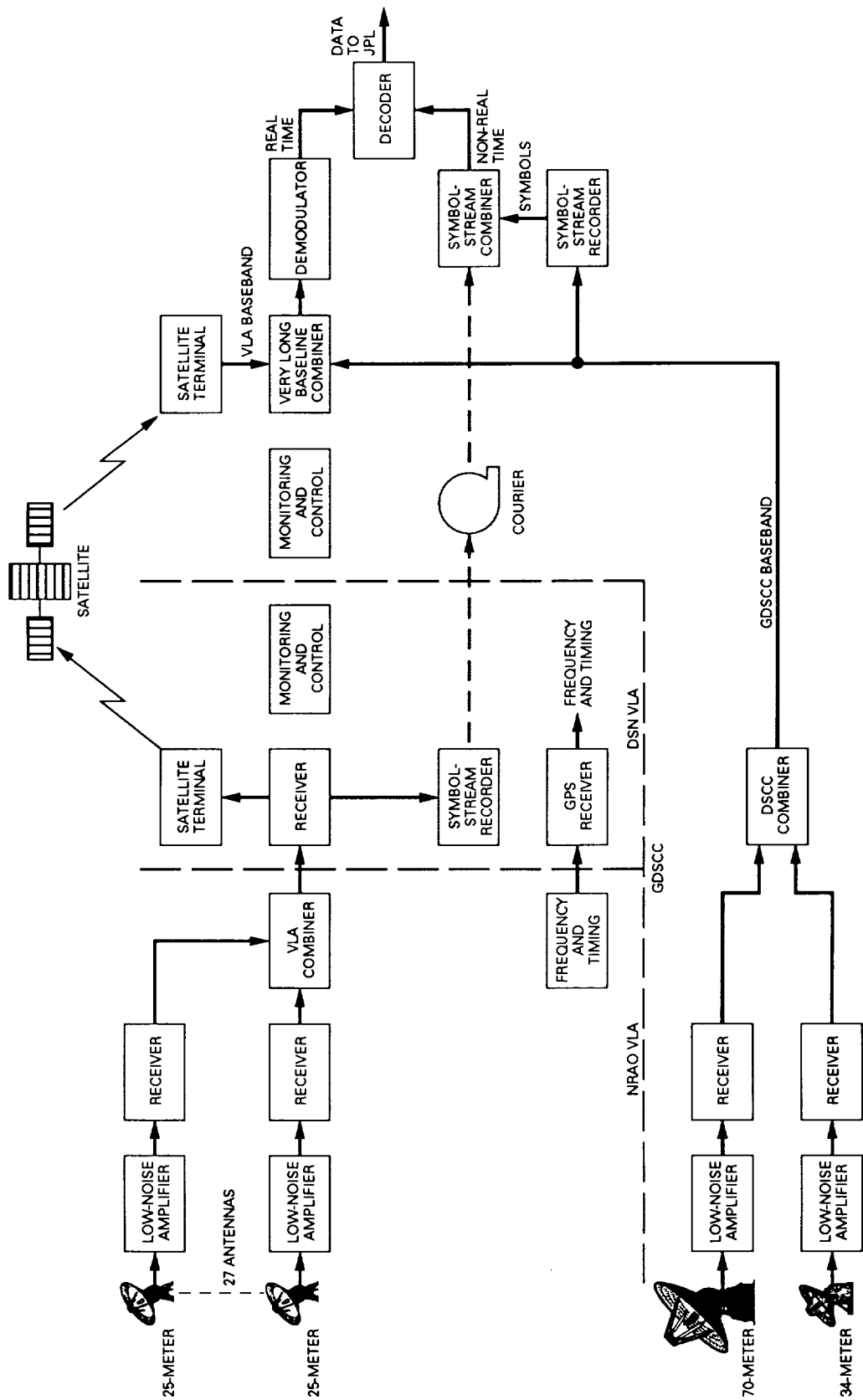


Fig. 1. The VGTA system.

**Table 4. VGTA bit-error-rate performance**

Nominal midpass conditions			
Data rate	7.2 kbps	14.4 kbps	21.6 kbps
SSNR	~6 dB	~5 dB	~3 dB
Real-time combining BER performance under nominal midpass conditions			
NDPP/MRT <sup>a</sup> (BER anomaly)	$10^{-5}$	$2(10^{-5})/2(10^{-4})^c$	$2 \times 10^{-4}$
CDT/NET <sup>b</sup> (gap compensation)	$<< 10^{-5}$	$2 \times 10^{-5}$	$3 \times 10^{-4}$
$\geq$ DOY 165 (timing fix)	N/A	$<< 10^{-5}$	$2 \times 10^{-5}$
Symbol-stream combining (non-real-time) BER performance under nominal midpass conditions			
MRT <sup>a</sup> (BER anomaly)	$<< 10^{-5}$	N/A	N/A
CDT/NET <sup>b</sup> (gap compensation)	N/A	$3 \times 10^{-5}$	$5 \times 10^{-4}$
$\geq$ DOY 165 (timing fix)	N/A	N/A	N/A

<sup>a</sup> NDPP/MRT = Neptune Dual Processor Program (Test)/Mission Readiness Tests.  
<sup>b</sup> CDT/NET = Configuration Demonstration Test/Near Encounter Test.  
<sup>c</sup> Greater than GDSCC stand-alone value during NDPP test; the MRT value is higher because of lower SNR.

**Table 2. Gapped bit error rates<sup>a</sup>**

Reed-Solomon (R-S) Coded Performance		
Condition	Array SSNR	R-S BER
VLA SNR $\ll$ GDSCC SNR	0.0 dB	$\ll 1 \times 10^{-5}$
VLA SNR $\geq$ GDSCC SNR	+0.5 dB	$\ll 1 \times 10^{-5}$
Convolutional Coded Performance		
Condition	Array SSNR	MCD <sup>b</sup> BER
VLA SNR $\ll$ GDSCC SNR	0.0 dB	$< 5 \times 10^{-3}$
	+1.5	$< 5 \times 10^{-5}$
VLA SNR = GDSCC SNR	+0.5 dB	$< 5 \times 10^{-3}$
	+3.0 <sup>c</sup>	$< 5 \times 10^{-5}$
VLA SNR $\gg$ GDSCC SNR	$> 0$ dB <sup>c</sup>	$< 2.5\%$ of GDSCC
VLA Stand-alone	$> 0$ dB	$< 1.5 \times 10^{-2}$

<sup>a</sup> *VGTA System Requirements and Design*, Table C-1, Document 1220-2 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 15, 1987.

<sup>b</sup> Maximum-Likelihood Convolutional Decoder.

<sup>c</sup> Domain of BER anomaly.

**Table 3. Midpass SSNR performance <sup>a</sup>**

Date	SSNR, dB					
	Site	Actual Performance	Design Performance	Site	Actual Performance	Design Performance
DOY 239	GDSCC	0.1	-0.9	CDSCC	0.7	-0.2
	VLA	2.3	1.0	Parkes	-3.7	-4.0
	VGTA	4.2	2.8	PCTA	1.8	1.3
DOY 240	GDSCC	0.3	-0.9	CDSCC	2.25	1.5
	VLA	2.1	1.0	Parkes	-2.0	-2.3
	VGTA	4.1	2.8	PCTA	3.35	3.0

<sup>a</sup> The data rate was 21.6 kbps, except for the CDSCC, Parkes, and the PCTA on DOY 240, when it was 14.4 kbps. The GDSCC and CDSCC used 70-m + 34-m HEF antennas.

**Table 1. Twenty years of telemetry arraying**

Mission	Date <sup>a</sup>	Location of antennas	Antenna types	Combiners <sup>c</sup>
Pioneer 8	1970	Spain	26-m + 26-m	Passive BB
Mariner Venus-Mercury (Mercury)	Sept. '74	U.S.	64-m + 2 (26-m)	R&D BB
Voyager 1 (Jupiter)	Mar. '79	U.S.	64-m + 34-m	RTC
Voyager 2 (Jupiter)	Jul. '79	U.S.	64-m + 34-m	RTC
Pioneer 11 (Saturn)	Sep. '79	U.S.	64-m + 34-m	RTC
Voyager 1 (Saturn)	Aug. '80	All DSCCs	64-m + 34-m	RTC
Voyager 2 (Saturn)	Aug. '81	All DSCCs	64-m + 34-m	RTC
International Cometary Explorer (Giacobini-Zinner)	Sep. '85	All DSCCs Spain/U.S. Australia/Usuda	64-m + 34-m <sup>b</sup> 64-m + 64-m 64-m + 64-m	Passive BB SSRC R&D SSRC R&D
Voyager 2 (Uranus)	Jan. '86	Spain U.S. Australia	64-m + 34-m 64-m + 2 (34-m) 64-m + 2 (34-m) + Parkes 64-m	RTC/BBA RTC/BBA RTC/BBA + LBC + SSRC R&D
Voyager 2 (Neptune)	Aug. '89	Spain U.S.  Australia	70-m + 2 (34-m) 70-m + 2 (34-m) + VLA 27 (25-m)  70-m + 2 (34-m) + Parkes 64-m	RTC/BBA RTC/BBA + VLBC + SSRC RTC/BBA + LBC + SSRC

<sup>a</sup> Listing of month and year indicates encounter period.

<sup>b</sup> The Spain and U.S. 64-m antennas also combined dual channels with passive BB.

<sup>c</sup> BB = Baseband (symbol-modulated subcarrier).

LBC = Long Baseline Combiner (baseband at ~300 km).

RTC = Real-Time Combiner, first version (baseband at ~30 km).

RTC/BBA = Operational RTC—part of BBA (Baseband Assembly).

SSRC = Symbol-Stream Recording and Combining (non-real time).

VLBC = Very Long Baseline Combiner (baseband at ~1,000 km).

- [24] W. J. Hurd, J. Rabkin, M. D. Russell, B. Siev, H. W. Cooper, T. O. Anderson, and P. U. Winter, "Antenna Arraying of Voyager Telemetry Signals by Symbol Stream Combining," *TDA Progress Report 42-86*, vol. April-June 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 131-142, August 15, 1986.
- [25] R. L. Horttor, D. W. Brown, and K. P. Bartos, "Readying the VLA for Voyager's Neptune Encounter: System Test and Performance," *JPL Highlights 1988*, JPL Publication 400-350, Jet Propulsion Laboratory, Pasadena, California, pp. 88-89, February 1989.

- [10] S. Weinreb, H. Dill, and R. Harris, *Low-Noise, 8.4 GHz, Cryogenic GASFET Front End*, VLBA Technical Report No. 1, National Radio Astronomy Observatory, Charlottesville, Virginia, August 29, 1984.
- [11] S. M. Petty, J. Bautista, and J. S. Ulvestad, "Readying the VLA for Voyager's Neptune Encounter: Cryogenic Amplifiers and Arraying Tests," *JPL Highlights 1987*, JPL Publication 400-329, Jet Propulsion Laboratory, Pasadena, California, pp. 32-34, November 1987.
- [12] W. D. Brundage, *NRAO VLA-GDSCC Telemetry Array Project, Quarterly Status Report*, National Radio Astronomy Observatory, Socorro, New Mexico, April-June 1987.
- [13] F. Manshadi, D. A. Bathker, and H. W. Marlin, "VLA Feedhorn for Voyager Encounter of Neptune," *TDA Progress Report 42-86*, vol. April-June 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 216-226, August 15, 1986.
- [14] J. S. Ulvestad, G. M. Resch, and W. D. Brundage, "X-Band System Performance of the Very Large Array," *TDA Progress Report 42-92*, vol. October-December 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 123-137, February 15, 1988.
- [15] J. S. Ulvestad, "Phasing the Antennas of the Very Large Array for Reception of Telemetry From Voyager 2 at Neptune Encounter," *TDA Progress Report 42-94*, vol. April-June 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 257-273, August 15, 1988.
- [16] S. J. Dolinar, "VLA Telemetry Performance with Concatenated Coding for Voyager at Neptune," *TDA Progress Report 42-95*, vol. July-September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 112-133, November 15, 1988.
- [17] W. D. Brundage, *NRAO VLA-GDSCC Telemetry Array Project, Quarterly Status Report*, National Radio Astronomy Observatory, Socorro, New Mexico, January-March 1988.
- [18] W. D. Brundage, *VLA-GDSCC Telemetry Array Project, VLA Reliability Review No. 1*, National Radio Astronomy Observatory, Socorro, New Mexico, May 2, 1986.
- [19] W. D. Brundage, *NRAO VLA-GDSCC Telemetry Array Project, Final Quarterly Report*, National Radio Astronomy Observatory, Socorro, New Mexico, July-October 1989.
- [20] W. D. Brundage, *VLA Implementation Plan, VLA-Goldstone Telemetry Array*, National Radio Astronomy Observatory, Socorro, New Mexico, September 29, 1987.
- [21] *NRAO Voyager Front-End Construction Plan, Revision A*, National Radio Astronomy Observatory, Charlottesville, Virginia, February 15, 1985.
- [22] H. W. Cooper and L. R. Hileman, "Software Development for the VLA-GDSCC Telemetry Array Project," *TDA Progress Report 42-87*, vol. July-September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 174-178, November 15, 1986.
- [23] T. A. Rebold and J. F. Weese, "Parkes Radio Science System Design and Testing for Voyager Neptune Encounter," *TDA Progress Report 42-99*, vol. July-September 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 189-205, November 15, 1989.

## Acknowledgments

The results described in this article were made possible by the dedication and effort of the many individuals of the engineering and operations teams at JPL and at the four arraying sites. Special notice is due the NRAO Central Development Laboratory for the low-noise amplifiers, providing two decibels added margin over the original baseline design for the VLA, and to Steve Howard of the Voyager Project for his invaluable assistance in analyzing the bit-error-rate performance of the VGTA throughout the simulation and test phase.

## References

- [1] J. W. Layland, A. M. Ruskin, D. A. Bathker, R. C. Rydgig, D. W. Brown, B. D. Madsen, R. C. Clauss, G. S. Levy, S. J. Kerridge, M. J. Klein, C. E. Kohlhasse, J. I. Molinder, R. D. Shaffer, and M. R. Traxler, "Interagency Array Study Report," *TDA Progress Report 42-74*, vol. April-June 1983, Jet Propulsion Laboratory, Pasadena, California, pp. 117-148, August 15, 1983.
- [2] D. W. Brown, H. W. Cooper, J. W. Armstrong, and S. S. Kent, "Parkes-CDSCC Telemetry Array: Equipment Design," *TDA Progress Report 42-85*, vol. January-March 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 85-110, May 15, 1986.
- [3] *An Introduction to the NRAO Very Large Array*, R. M. Hjellming, ed., National Radio Astronomy Observatory, Socorro, New Mexico, April 1983.
- [4] D. L. Trowbridge, J. R. Loreman, T. J. Brunzie, and R. Quinn, "An 8.4-GHz Dual-Maser Front-End System for Parkes Reimplementation," *TDA Progress Report 42-100*, vol. October-December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 301-319, February 15, 1990.
- [5] R. Stevens, "Applications of Telemetry Arraying in the DSN," *TDA Progress Report 42-72*, vol. October-December 1982, Jet Propulsion Laboratory, Pasadena, California, pp. 78-82, February 15, 1983.
- [6] L. J. Deutsch, "An Update on the Use of the VLA for Telemetry Reception," *TDA Progress Report 42-72*, vol. October-December 1982, Jet Propulsion Laboratory, Pasadena, California, pp. 51-60, February 15, 1983.
- [7] J. W. Layland, P. J. Napier, and A. R. Thompson, "A VLA Experiment — Planning for Voyager at Neptune," *TDA Progress Report 42-82*, vol. April-June 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 136-142, August 15, 1985.
- [8] J. W. Layland and D. W. Brown, "Planning for VLA/DSN Arrayed Support to the Voyager at Neptune," *TDA Progress Report 42-82*, vol. April-June 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 125-135, August 15, 1985.
- [9] K. Bartos and W. D. Brundage, "The Voyager 2 Neptune Encounter, A Management Challenge," *PM Network*, vol. III, no. 4, Project Management Institute, Webster, North Carolina, pp. 7-21, May 1989.

### C. Bit-Error-Rate Performance

Table 4 lists the bit-error-rate performance for the VGTA under several typical conditions and as a function of time as the system was optimized. (The successive improvements were described under the "VGTA Test and Operations" section.) The domain<sup>36</sup> of the anomalous BER is noted in Table 2, which summarizes the design expectation. Final performance met or exceeded the expectation in all respects.<sup>37</sup>

The symbol-stream performance was initially equivalent to that achieved with the gap power-level compensation of the real-time system. The implication here was that the symbol-combiner-to-decoder interface level was more gap-tolerant than the equivalent real-time path, where the demodulator-synchronizer intervened. Performance of the symbol-stream mode was expected to be improved at high data rates with the VLA timing fix, as was the real-time data mode. Once the encounter period began, the real-time reliability was so high that no playbacks were fully processed. Data for this period are indicated "N/A" in Table 4.

PCTA bit-error-rate performance was statistically equivalent to the prevailing bit SNR (SSNR + 3 dB) and therefore was not monitored in detail, other than through the routine operation of the Voyager Project correction count plotters (Figs. 2 and 3).

### D. Imaging

Figures 6 through 9 illustrate the final performance of the VGTA at 21.6 kbps, the most sensitive rate, both in margin and format. Figure 6 is the fully encoded real-

time image of Neptune received about midpass on the day before encounter. As expected with the full Reed-Solomon coding, the image is error-free. The convolutional bit error rate of  $3.7 \times 10^{-5}$  at the time of the image is somewhat greater than the  $2 \times 10^{-5}$  average value of Table 4. It is well within the typical variance of the decoding process.

Figure 7 is a raw playback image of Triton, taken early in the encounter pass and played back two days later, where the elevation angle effects at Goldstone, together with the VLA data gap, are degrading the bit error rate to  $1.7 \times 10^{-4}$ . In the original glossy photograph, perhaps two dozen errors are seen in the dark part of the field. Figure 8 is the same image after "de-spike" processing, a smoothing algorithm used in the Multimission Image Processing System. No hits are visible, and the apparent resolution is undegraded. These two figures show that the system performed in the worst-case mode (convolutional coding only) with minimal visual bit-error impact. Comparable images taken before the fixes indicated in Table 4 show considerable visual degradation. And finally, Fig. 9, showing a segment of Triton, is again a raw playback image, taken at about midpass. A few hits are visible.

## VII. Conclusion

The successful implementation and operation of the interagency telemetry arrays permitted Voyager 2 spacecraft operation at 21.6 kbps for the full view period when Voyager was over Goldstone and Canberra, with the usual DSN 90-percent weather confidence (approximately 2 dB margin) as well as weather diversity for more catastrophic conditions, at all data rates. Together with the expansion of the 64-m antennas to 70 meters, interagency arraying doubled the science-data return from the Voyager-Neptune encounter.

It was further demonstrated that, for a special event, the resources of diverse agencies could be melded successfully and with high operational reliability.

<sup>36</sup>High data rates and SNRs, with significant VLA contribution.

<sup>37</sup>S. Howard, "GDSE Final Test Report for the Near Encounter Test (NET): Final Evaluation of Interagency Array Support," JPL IOM Voyager-GDSE-89-033 (internal document), Jet Propulsion Laboratory, Pasadena, California, July 27, 1989.

link indicated losses of 0.1 dB or less, within the absolute uncertainty of the estimators. Observations of Voyager made at the VLA (independent of telemetry performance) during the early test phase would corroborate a stronger Voyager signal than predicted.

Subtracting this inferred 0.5 dB from the average Parkes residual of +0.3 dB suggests a shortfall in Parkes G/T of 0.2 dB, which is consistent with a reported typical system temperature of 23.5 K versus the design value of 22 K.<sup>33</sup> The reported antenna gain of 71.6 dBi at a 45-degree elevation angle is consistent with the baseline design value of 71.2 dBi at the Voyager elevation angle.

Similar reasoning suggests G/T actuals of +0.3 dB for CDSCC (70-m antenna + 34-m HEF antenna) and of +0.6 dB for GDSCC, as compared with the respective design values.<sup>34,35</sup> These deltas are partially explained by the known pessimism in the gain values for the 34-m HEF antennas in the referenced documents. A further consideration in this analysis may be differing star-calibration references used by the VLA and the DSN.

The typical indicated combining loss was 0.2 dB, which was the design goal but which was subject to the uncertainty of three simultaneous estimators. Losses of 0.1 dB or less were observed under closely controlled preshipment tests at JPL.

## B. Weather Diversity

The estimated SSNR of the combined signal at each complex was made available in real time to the Voyager Project, where it was plotted in conjunction with the bit error count as derived from the Reed-Solomon error-correction process (see Figs. 2 and 3). The bit error rate can be approximated by dividing the count by the bit rate times 48, the averaging time in seconds.

One of the virtues of arraying geographically separated sites is the protection offered against inclement weather. Figure 2 shows the performance for DOY 222 for most of the U.S. and Australian viewing period. The pass began with severe weather-caused degradation at Goldstone, such that for approximately two hours the VLA stand-alone signal was processed at the data rate of 14.4 kbps. The saturated correction count in that period reflects the

$1.5 \times 10^{-2}$  error rate of the gapped VLA data stream. As this signal format was fully coded, no data degradation was suffered. The full array of VLA + 70-m antenna + 34-m antenna was resumed for most of the pass after the weather cleared.

Shortly before the handover to Australia, the spacecraft sequenced to a data rate of 21.6 kbps, with expected PCTA performance. About three hours later, rain at Tidbinbilla took the system below the SSNR threshold of 0 dB, yielding uncorrected bit errors in the playback data as well as exceeding the Reed-Solomon error-correction threshold. Unfortunately, the Parkes contribution was insufficient to maintain adequate margin.

A few days later, as shown in Fig. 3, margins were maintained throughout, despite very severe weather at the VLA. The rise of correction rates at both ends of the Goldstone pass reflects the effect of the low elevation angle on the Goldstone SNR during the VLA data gaps in the manner expected. The SSNR shows a higher short-term variance than appears in Fig. 2, owing to the shorter integration times of the estimates. Figure 4 repeats the first half of Fig. 3(b), showing in more detail the constituent SSNRs. With a threshold of 0 dB, the Goldstone array (bottom trace) could have supported reception alone for about three hours (during midpass). On the other hand, the VLA alone would have supported real-time imaging for all but about one hour, although the unprotected playback data would have been severely impacted. With Goldstone and the VLA working together, support was 100 percent throughout, as evidenced by the low correction count of Fig. 3(a).

Figure 4 is taken from plotted data generated at Goldstone in real time. To provide visibility into the arraying operation, such plotting capability was implemented for each combining site. Four-color plots were generated, indicating the estimated SSNR of each complex and of the combined data stream, as well as the computed ideal sum of individual streams. Because the plots were updated several times per minute, they provided operations personnel with an immediate visual indication of the arraying operations' status. They also served, as indicated by Fig. 4, as an efficient means of after-the-fact engineering analysis.

Finally, Fig. 5 plots the VGTA performance on encounter day, with the several data rate changes. Weather and performance were flawless. By comparing the several traces, it was readily possible to distinguish between site anomalies and spacecraft maneuvers, which occurred at about 05:40 and 06:20 hours UT (Earth receive time).

<sup>33</sup> PCTA System Requirements and Design, Update for Neptune Encounter, Document 1200-2A (internal document), Jet Propulsion Laboratory, Pasadena, California, July 15, 1988.

<sup>34</sup> Ibid. [Reference is to entire document.]

<sup>35</sup> VGTA System Requirements and Design, Document 1220-2 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 15, 1987.

bit too late. The 280  $\mu$ sec or so of full noise, but no coherent VLA signal, was being bridged at the lower rates, but not at 14.4 kbps and higher. As was the case with the dropouts (discussed above), this had been a negligible SNR effect for astronomy. The NRAO proposed a succinct plan to fix the problem, complete with fallback plans to be used in the event of an unforeseen problem. By mid-June, with the fix,<sup>30</sup> the 21.6-kbps BER was as modeled. The playback errors would be minimal, as initially expected!

As noted above, VLA and DSN operations personnel had been heavily involved in the final stages of implementation and testing. Operations staffing consisted of two dedicated DSN personnel at each site, in addition to shift operators, all of whom were experienced in VGTA operations prior to the encounter phase. During critical operations, supporting specialists were available at the VLA and at Goldstone on short notice. Operations planning and other considerations, including intersite voice and predict transmission, are discussed elsewhere.<sup>31</sup>

From January through May 1989, a number of mission-readiness tests were performed to check the overall system preparation and the operating procedures. These tests also constituted the final training opportunities for the operations personnel in all their areas of responsibility, including items such as preparing the configuration files that controlled the VLA, monitoring the VLA performance, maintaining the operational status of the JPL equipment, transmitting data to Goldstone for combining, and making backup recordings of the data stream. These tests were successful and led naturally to delivery of the first committed Voyager data late in the spring of 1989.

A tribute to the thoroughness of the system testing is that it uncovered shortcomings early enough to enable corrective actions far enough ahead of the critical tracking periods to provide the confidence and the assurance that the data reception during encounter would be successful.

## VI. Telemetry Performance

Three fundamental measures of performance used to characterize the arrays were

- (1) front-end G/T, referenced to known radio sources
- (2) symbol signal-to-noise ratio (SSNR)

- (3) convolutional decoding bit error rate, measured as the error-correction rate in the Reed-Solomon decoding process

Two other measures were the estimation of the carrier-to-noise-density ratio and of the bit signal-to-noise ratio. These estimations were considered secondary, especially in the VGTA case, wherein the gapped signal caused errors of as much as one dB and more.

The front-end G/T was discussed in Section III; the following paragraphs summarize the telemetry performance.

### A. SSNR Performance

The first function of the symbol-stream recorders (see Fig. 1) at each site was baseband demodulation to detect symbols for recording. The demodulators also measured SSNR by means of "split-symbol moment estimation." In addition to serving the recording function, these demodulators thus provided the primary real-time performance data type. Similar units were a part of the standard processing stream and therefore provided a measure of the combined SSNR. The accuracy of the estimates was derived from calibration (at CTA 21) against the symbol error rate of a known sequence. With averaging, this accuracy, over the range of interest, was 0.1 to 0.2 dB. Table 3 presents typical performance at midpass for both arrays on two different days of the year (DOY) in 1989. As indicated, both arrays typically employed the 34-m high-efficiency (HEF) antennas at each DSCC; they occasionally included the 34-m standards, whose contribution was normally not required.

The VLA's observed performance typically exceeded that set out in the baseline design<sup>32</sup> by the factor of

$$\left[ \frac{62}{60} \text{ (efficiency)} \right] \times \left[ \frac{35}{32} \text{ (temperature)} \right] \times \left[ \frac{26}{25} \text{ (available antennas)} \right] = 0.7 \text{ dB}$$

which, subtracted from the average residual in Table 3 of +1.2 dB, yields an inferred positive residual of 0.5 dB for the received Voyager signal, assuming that the pointing loss, combining loss, and receiver losses were nominal: 0.2, 0.1, and 0.4 dB, respectively. Calibrations of the intersite

<sup>30</sup> A simple wiring change at each antenna.

<sup>31</sup> *Operations Planning for the VLA-GDSCC Telemetry Array*, Document 1220-8 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 1988.

<sup>32</sup> *VGTA System Requirements and Design*, Document 1220-2 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 15, 1987.

Solomon error-correction threshold as regards the partially gapped data stream. The model and simulations were verified.<sup>25</sup>

After months of intensive strip-chart and triggered scope monitoring of the VLA signal output and of the detected carrier, the cause of the dominant dropout was diagnosed and fixed in September 1988 by a single character correction in the operating-system code. A number of other, smaller spikes and keyboard-induced anomalies continued and were gradually abated before committed Voyager operations began.

Installation of the majority of the JPL equipment at the VLA and at Goldstone occurred in 1988. As each of the major assemblies was completed, it was subjected to system testing in the engineering lab and the CTA 21 test facility before being shipped to its final destination for integration and on-site system testing, as outlined in the *VGTA System Test Requirements* document.<sup>26</sup> These tests included validation of most of the functions that had been exercised in the monthly test periods at the VLA, as well as validation of the complete Goldstone arraying configuration.

Monthly tests done late in 1988 continued with a variety of goals. Training of DSN operations personnel began, and procedures were developed for use during the encounter period. Software features of the new VLA computer system, including system status displays designed specifically for JPL use,<sup>27</sup> were iterated monthly. Observations of the spacecraft were made on a regular basis. Tests frequently involved varying the number of active antennas being combined so that typical encounter SNRs could be simulated with the prevailing data rate and understanding of the system throughput could be confirmed. Tests of radio-frequency interference caused by JPL equipment, particularly the satellite transmitter that shipped data to Goldstone, were made to determine the level of caution that was needed during normal astronomical observing.

As 1989 began, there were still unfinished VLA software tasks and several continuing anomalies. As planned in 1987, a suite of tests was developed and performed on a

monthly basis; these tests followed the regular VLA software updates. Since freezing the VLA system for nine months was an unacceptable option, this set of tests served to verify that performance was proper in all items relevant to Voyager tracking. Relatively minor software changes were made for JPL purposes, in addition to other changes for astronomical use prior to the three-month software freeze for the encounter period.

One of the most significant continuing problems was the "anomalous BER,"<sup>28</sup> which occurred at high SNRs and had convolutional bit error rates (BERs) an order of magnitude or two higher than the model and early simulations would indicate.<sup>29</sup> As noted above, this would not degrade the Reed-Solomon threshold significantly, but the playback portion of the 21.6-kbps frame would be hit with more errors than expected. Immediately following the NDPP test, Voyager analysts made the discovery that all the errors were occurring during a gap—furthermore, at the end of the gap. Because the early simulations were performed at somewhat lower SNRs, which were nearer the system design point, it was not clear whether or not the problem might have been overlooked in the simulations. Consequently, while the system was being otherwise finalized, simulations were renewed at both CTA 21 and Goldstone, concentrating on the effect of the drop in total power level (as well as in SNR), during the gap, upon the DSN telemetry string.

By early 1989, a gap-triggered gain changer (gap power-level compensator), located between the baseband combiner and the decoder, was producing ideal performance at 7.2 kbps—again, the highest currently available rate. The opportunity to check performance at 14.4 and 21.6 kbps was awaited with high expectations as the encounter phase approached. But, alas, the compensation was not very effective at rates above 9.6 kbps. The system would fully meet requirements in the predominant Reed-Solomon mode, while the playbacks, although not meeting full expectation, would be processable and scientifically useful. Final acceptance of the VGTA system was based upon successful tracking of the Voyager 2 spacecraft, combining of the baseband and symbol-stream data at Goldstone, and delivery of the data to the Voyager Project, thus demonstrating that performance was within specifications.

Continued vigilance with the strip-chart recorder paid off spectacularly in May. One small oddity led to the realization of a timing problem immediately following the gap. The antenna phase switchers were being turned back on a

<sup>25</sup> S. Howard, *Evaluation of VLA-Goldstone Telemetry Array Performance During the NDPP Test, 1988*, Document D-6112 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 17, 1988.

<sup>26</sup> *VGTA System Test Requirements*, Document 1220-6 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 1, 1987.

<sup>27</sup> *VGTA System Requirements and Design*, Document 1220-2 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 15, 1987.

<sup>28</sup> First noted during the NDPP test at 14.4 kbps.

<sup>29</sup> S. Howard, *op cit*.

Because of the high reliability requirements imposed on the equipment, redundant recording channels were implemented at all four sites. The option to record both the local and remote signals at Goldstone and at Canberra was designed into the assemblies. The ability to play back and combine existed at all sites, thus permitting self-test at each site as well as operational combining at Goldstone and at Canberra. This test feature was used extensively during test and training periods.

To record symbols, standard DSN demodulator-synchronizer assemblies were incorporated into the design so as to remove the subcarrier. This also provided the ability to make symbol SNR estimates at the VLA and at Parkes as well as estimates of the baseband combiner inputs and output, which gave a measure of the real-time system performance. To assure valid recordings, the symbol-stream equipment also made SNR estimations during recording, as well as upon playback and combining.

While limited system test time did not permit resolution of some operability problems, bit-error-rate performance met expectations (see Section VI). Very little operational experience was gained, as the excellent reliability of the real-time system obviated the need (and hence the opportunity) to exercise fully the backup system during the encounter phase. In any case, the benefits and advantages of symbol-stream recording were well demonstrated.

## V. VGTA Test and Operations

Monthly tests were performed at the VLA from mid-1985 through mid-1989. As described earlier, these tests concentrated on radio-frequency performance and array phasing in the period from 1985 through 1987. Late in 1987, these tests began to include more examination of the VLA back-end system and the DSN telemetry equipment. The term "back end" refers generally to the central control room electronics (intermediate-frequency channels, correlator, summers) and the on-line computer system (antenna control, autophasing, monitoring).

Following the Voyager-Uranus encounter in 1986, some of the Parkes arraying equipment was temporarily installed at the VLA, together with a DSN convolutional decoder and recording equipment, to get a sample of the "real" data gap to confirm the simulations described earlier. By late spring of 1987, recordings were made of the decoded Voyager data stream at 7.2 kbps, the highest rate available during the cruise phase. These recordings were returned to the Voyager Project for analysis. Since intersite communications and Goldstone combining were not available

at that time, the stand-alone data stream was evaluated because it was the best available first look at the true gap effect. Results were comparable to the modeled  $1.5 \times 10^{-2}$  error rate. The first look at combined data and at the encounter data transmitted at rates of 14.4 and 21.6 kbps would await the Voyager Neptune Dual Processor Program (NDPP) tests.

Late in 1987, as part of an ongoing NRAO activity, the VLA on-line computer system was upgraded. The new computers, which were faster and had more memory, provided more flexibility and power in a variety of tasks for controlling and gathering data from the array. They also had none of the crashes that were common to the old computers, which had significant communication problems among a number of processing units. However, as with any new system, there were significant hurdles to be overcome. Much of the test time during the first half of 1988 was devoted to finding bugs in the system and checking new features that had been added.

Perhaps the most significant bug encountered was the signal "dropout," first observed in December 1987 with the new telemetry receiver and the new on-line system. These outages showed a variety of characteristics, but typically lasted on the order of one second. The typical occurrence rate was every 10 to 15 minutes. Although such glitches are tolerable for radio astronomers since they cause only a slight degradation in SNR, they would cause large gaps in telemetry streams.

Before the operating-system bug causing these dropouts was found, the Voyager NDPP tests were supported at the VLA in July 1988. These tests involved five Voyager tracks in which both the VLA and Goldstone participated, both to support checkout of the spacecraft encounter mode and to perform a preliminary evaluation of the VLA-Goldstone array. Although the dropouts were a major embarrassment during the NDPP test, the test objectives were met by minimizing the number of tasks to be performed by the operating system, thus keeping the signal outages to a tolerable level.

While the overall arraying system was not as yet in a mature state, the NDPP tests served to validate several key performance factors, including the effective  $G/\Gamma$  (in telemetry terms) of the then-existing 20 VLA antennas at 8.4 GHz,<sup>24</sup> the quality of the satellite data link, and the real-time signal combining at Goldstone [25]. A specific objective of one test period was the validation of the Reed-

<sup>24</sup> The antennas were outfitted at the average rate of nine per year from 1986 to the end of 1988.

for an outboard static-delay unit in the VGTA Very Long Baseline Combiner.

As related elsewhere in this article, power-level compensation during the VLA gap was required to optimize the bit error rate. Since the output data power level of the combiner resulted from summation of the VLA and Goldstone signals, the drop in data level during the gap depended upon the actual signal-to-noise ratio difference between the two complexes. Compensation was added immediately following the combiner and was synchronized to the gapped waveform. Several values of gain were provided. The level was set for a compromise between a constant noise power level for the demodulator-synchronizer and a constant signal power level for the decoder.

To meet reliability requirements, a backup combiner and compensator were operated in parallel with the operational equipment. In the event of problems with the on-line combiner, the backup combiner could be selected from the control terminal through appropriate commands.

### E. Test-Signal Generation

For the Uranus encounter, the test-signal generator (TSG) provided test signals at baseband and at 315 MHz. This made testing possible at the inputs to the receiver, the combiner, and the recording equipment as well as modulation of the X-band signal generator at Parkes for testing the complete system. The capability to add noise to the appropriate signals was incorporated into the TSG so that testing could be done at anticipated signal-to-noise levels. Ranges of SNR and power were designed to simulate encounter levels and provide  $\pm 10$ -dB margins. Having two baseband channels permitted dynamic simulation of the differential signal delay in combiner testing.

For the Neptune encounter, the TSG was modified to provide baseband signals only, with associated noise. The 315-MHz and the 18.75-MHz signal-generation capability was built into the receivers, along with a minimal modulation capability, to permit stand-alone testing of the receivers.

A major change in the test-signal generation came about because of the use of the satellite link between the VLA and Goldstone. No longer was it practical to send the test signal from Goldstone to the VLA, inject it into the receiver, and then return the demodulated baseband back to Goldstone for combining, data reduction, and performance evaluation; this would require the use of a two-way satellite link. To save such costs, the TSGs at Goldstone and Canberra were modified to be synchronized to those

at the VLA and at Parkes, respectively, through the use of 1 pulse per second (pps) timing signals derived from stable frequency and timing standards at all four sites. Also incorporated into the system design was the ability to measure the time delay between sites through the use of the 1-pps signals, as was required for baseband combiner initialization. Knowledge of intersite timing offsets was maintained through the Global Positioning System (GPS). The requirement of 1  $\mu$ sec, to facilitate tape playback alignment, was readily met. An added feature was the ability to generate convolutionally coded data so that the VGTA/PCTA equipment could conduct self-test through the convolutional decoders.

To further simulate the signal received at the VLA, the test-signal generator in the VLA receiver was capable of creating the 1.6-msec gap in the data. This capability made possible testing of the gap effect on the performance of the receiver, the baseband combiner, and the symbol-stream recording and combining. A gap simulator was also provided at Goldstone to allow gap performance testing on a stand-alone basis. This capability proved very valuable during the evaluation of the gap power-level compensation.

### F. Symbol-Stream Recording and Combining

The backup recording identified in Fig. 1 was deemed necessary to achieve the required reliability of 98 percent for the added equipment at each site, including the communication link, by permitting post-pass playback of the tapes at Goldstone or Canberra. During the Uranus encounter, PCTA baseband recordings were made by utilizing modified Mark III Data Acquisition Terminals (DATs). Use of the DATs was justified largely on the premise that most radio astronomy facilities already had such devices, which could be simply and quickly accessed for arraying purposes without impairing VLBI usage.

In the meantime, development work was under way at JPL to demonstrate the feasibility of symbol-stream recording and combining. This approach offered several advantages, including a lower recording density requirement (because the subcarrier had been removed), permitting the use of standard computer-compatible tape. The equipment would also prove to be simpler and more reliable than the complex adaptations necessary to utilize the DATs. By 1986, a demonstration of the symbol-stream capability had been accomplished [24] and a plan developed to utilize this technique for both the VGTA and PCTA at Neptune encounter.<sup>23</sup>

<sup>23</sup> H. Cooper, *VGTA/PCTA Symbol-Stream Recording and Combining Subsystem (SSRC) Plan* (internal document), Jet Propulsion Laboratory, Pasadena, California, June 2, 1986.

routines that vectored the interrupt to the desired "C" routine and then performed the required end-of-interrupt operations. Several other problems were solved by using special utility programs [22]. A large benefit of the "C" language was that it resulted in a 20-percent decrease in code size with a 50-percent increase in speed.

## B. PCTA X-Band Front End

Arrangements were made with the European Space Agency (ESA) to borrow the dual-maser X-band front end that had been shared at Parkes for the Voyager-Uranus and Giotto-Comet Halley encounters. This consisted of a CSIRO-designed feedhorn, JPL-supplied waveguide assemblies, and ESA-procured traveling-wave masers of JPL design. Reimplementation for the Neptune encounter included an extensive refurbishment of the maser amplifiers, and new monitor and control equipment based upon standard DSN designs [4]. Custom features included remote monitoring of cryogenic functions to Tidbinbilla to facilitate maintenance during unattended periods and a noise-adding radiometer for antenna-pointing calibrations.

The Parkes telemetry implementation was coordinated with the Radio Science open-loop recording installation, sharing facilities and the X-band front end [23].

## C. Telemetry Receiver

The same basic receiver that was used for the Uranus encounter was also used for Neptune. Owing to the increased space loss (3.5 dB), the increased Doppler velocity change (350 kHz versus 120 kHz), and the increased acceleration (75 Hz/sec versus 9 Hz/sec) at encounter, some design changes had to be made in the receiver to accommodate the received signal.

The receiver was originally designed to operate at the Parkes facility in Australia, receiving an input signal at about 315 MHz that had been down-converted from X-band. This design was again employed for use at Parkes; however, at the VLA, the interface from the VLA receiver to the DSN receiver was made at a frequency of about 18.75 MHz. To accommodate this interface, an up-converter was designed to accept the 18.75-MHz signal and provide a signal at 315 MHz for input to the receiver.

The VLA was fixed-tuned for each pass at a frequency placing the received spectrum within special 8-MHz band-pass filters within the available 0-MHz to 50-MHz pass-band. The frequency of 18.75 MHz was chosen with consideration for aliasing of the VLA sampling response as well as of the following up-conversion process.

The acceleration rate of 0.9 Hz per second at Uranus allowed the existing 10.8-Hz loop bandwidth to be used. A bandwidth of 21.6 Hz was added for use at the Neptune encounter. This bandwidth provided the best compromise between loop SNR and acceleration capability. The increased Doppler change during encounter made it mandatory to program the first local oscillator of the receiver. This was done by inputting discrete ramps into the local-oscillator synthesizer. The ramp rate was determined by sampling the phase error in the loop and then applying the ramp to maintain this error at less than a preset amount.

For the Uranus encounter, predicted frequencies were input in the standard DSN format.<sup>22</sup> Since that format bore little relation to the Parkes receiver configuration and because of the added operational needs at the VLA, requirements were placed on the DSN Network Operations Control Center (NOCC) Support Subsystem to generate predictions for all interagency sites at the X-band received "sky" frequency. The VGTA/PCTA receiver acquisition design was modified to operate principally in this mode. Other options were provided, including the use of actual VLA/Parkes interface frequencies. These changes greatly facilitated operating procedures and improved acquisition times. Given any of the input options, the receiver would automatically acquire the spacecraft signal upon command.

## D. Very Long Baseline Combiner/Long Baseline Combiner

The function of the baseband combiners included

- (1) ephemeris-driven differential delay to time-align the two baseband signals
- (2) cross-correlation of the two to measure the residual error in alignment
- (3) loop closure to track out this error
- (4) summation of the two data streams, after appropriate weighting according to the prevailing input SNRs

The combiners for Neptune were basically the same as at Canberra for the Uranus encounter, with the addition of more static delay at Goldstone to take into account the satellite-link delay from the VLA, an increase to four times the dynamic delay range for the longer VGTA baseline, and adjustment of the loop parameters, primarily to accommodate the lower input SNRs of the PCTA. The combiners for the two arrays were essentially identical but

---

<sup>22</sup> Track synthesizer frequency (TSF) and Doppler frequency.

Throughout the implementation, system reliability was given special attention. Not only were the existing and the new equipment scrutinized, but spares, mean time to recovery, and operational procedures were studied. It was not at all clear that some 30 antennas could be confidently and repeatedly sustained and operated, given the diverse geographical and organizational aspects of the array. NRAO studies [18] identified areas in need of additional sparing; the NRAO also planned modes of operation so as to achieve maximal redundancy. For example, all four VLA intermediate-frequency channels were included in the Voyager-encounter design to support spacecraft failure modes resulting in left-circular polarization and to provide a spare channel for each polarization of the 27 antennas—108 channels.

In aiming for a 98-percent availability of the VLA common system, the provision of on-site primary power and the additional lightning protection described above were critical factors. Examples of sparing actions were the provision of a spare on-line computer to support the multi-computer upgrade that the NRAO was concurrently (1986–87) undertaking and of a JPL-supplied “hot backup” rubidium frequency standard. Redundant on-line equipment was provided for virtually all the Goldstone array equipment as well as for all of the telemetry equipment at the VLA, with the exception of the receiver, which was fully spared and of a well-proven design.

All of these preparations, as well as the proficiency of the NRAO and DSN operations personnel at the VLA and Goldstone and of the special support personnel standing by at both sites, paid off handsomely. Several measures of the availability of the VGTA and PCTA arrays reported by the DSN were all in excess of 99 percent of the data reaching the ground. As for the VLA signal to the DSN interface, for the 40 passes from April 26 through September 28, 1989, the availability was 99.959 percent, with 25 antennas available 99.989 percent of the time [19].

## IV. Implementation

The JPL implementation manager, functioning as a task manager in the line organization, had budgeting, scheduling, and reporting responsibilities for both the VGTA and PCTA implementations of the JPL subsystems. This included planning and scheduling coordination with both the NRAO and CDSCC, as well as project engineering functions for the VLA facilities augmentations in which JPL was involved. The overall VGTA implemen-

tation was covered in one internal document,<sup>19</sup> with the JPL subsystems detailed in another.<sup>20</sup>

For the VGTA, the VLA-Voyager preparation manager and the Central Development Laboratory (CDL) manager had responsibility for work undertaken by the NRAO. Implementation by the NRAO included design, fabrication, scheduling, installation, testing, training, and operations according to the *Management Plan for the VLA-GDSCC Telemetry Array Project*<sup>21</sup> and the *VLA Implementation Plan, VLA-Goldstone Telemetry Array* [20]. The CDL plan for the low-noise front ends is given in [21]. Installation into the VLA system was achieved with minimal disruption to mechanical and electronic systems, maintenance, upgrades, and observing schedules. The VLA observing schedules included normal testing of X-band as a standard VLA observing band, and periodic VGTA performance testing, as well as testing of the operational telemetry reception from Voyager.

The principal design changes for Neptune encounter, applicable to both the VGTA and PCTA, are briefly outlined below as an update to [2].

### A. JPL Software

The PCTA Uranus-encounter implementation used integral microprocessor computers to perform all of its control and monitor functions. While the assemblies all ran ROM-based firmware, the array controller contained floppy disk drives that ran the operating system. The PCTA software, written in Pascal, was developed in the array controller. While this software/firmware system worked, it had many difficulties. The Pascal compiler errors were numerous, and much time was spent finding workarounds for them.

Since much of the software and firmware had to be modified for the VGTA Project, IBM PCs were employed as the control terminals to enable Microsoft “C” to be used as the programming language. Several problems were successfully solved so that the IBM PC/“C” language could be used in conjunction with the 8086 multibus system used in most of the hardware assemblies. Real-time interrupts were handled through the use of simple assembly language

<sup>19</sup> *Project Implementation Plan for the VLA-GDSCC Telemetry Array*, Document 1220-7 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 15, 1987.

<sup>20</sup> H. Cooper, *VGTA/PCTA Implementation Plan* (internal document), Jet Propulsion Laboratory, Pasadena, California, January 20, 1987.

<sup>21</sup> *Management Plan for the VLA-GDSCC Telemetry Array Project*, Document 1220-1 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 15, 1985.

## F. VLA Facilities Augmentation

Electrical power reliability and the addition of real-time communications capability to transmit the data to the Goldstone Communications Complex were facilities changes anticipated in the VGTA Management Plan. Deterioration of the on-site power-distribution cables, as well as the need to increase the level of lightning protection, became additional challenges for the JPL-NRAO management team that were not envisioned at the outset. Each issue was resolved and the facilities were augmented to provide the level of reliability and service necessary to support the required availability figure.

VLA site personnel reported experiencing commercial power outages and fluctuations during summer thunderstorms or high-wind conditions associated with winter storms. The VLA site was instrumented in 1985 to determine the extent of the outages. With only one year of data, it became clear that the commercial power regularly dropped below 100 volts, a condition below the operating limits of the JPL electronic equipment to be used during encounter. Weather data correlated well with the frequent voltage transients and outages that occurred during inclement weather conditions.

DSN-standard power quality was required at the VLA site to support the encounter. An operational power system was located in Idaho and purchased in 1987 for the VLA application. The installation at the VLA was designed and installation completed in 1988 as a cooperative effort between the VLA facilities organization and JPL's Ground Antennas and Facilities Engineering Section.<sup>15,16</sup> The shortcomings of the existing commercial power system were overcome by installing two diesel-engine-driven generators, each rated to deliver 1,400 kW. Although one generator alone can normally supply the total power required, the two generators operated in parallel, each at 50 percent of rating. If a problem should cause automatic shutdown of either of the engine generators, the remaining generator would be fully capable of carrying the total power load of the site. This is the established mode of operation throughout the DSN, and it is referred to as "spinning reserve" mode.

The insulation on the direct burial cables that distribute power throughout the VLA site was gradually de-

teriorating, precipitating shorts to ground. Depending on the short location, as much as a whole arm of the "Y" could be taken out of service until the short was located and the faulty cable section was replaced. The one to two days to repair the fault were clearly in excess of the allowable mean-time-to-repair goal supporting the required availability figure. Special funding was obtained to replace the aging cables to the limit of "Y" utilization by the Voyager support configuration [17]. Replacement was completed in 1988.

Lightning struck on July 5, 1988! Evidence indicated that lightning entered the control building through an unprotected whip antenna, damaging several computers and disrupting telephone service. The adequacy of lightning protection was reviewed by an JPL/NRAO team, representatives from the Langmuir Laboratory for Atmospheric Research, and the Kennedy Space Center Lightning Safety Committee.<sup>17</sup> In addition to existing lightning rods on the control building, a lightning air terminal system above both the control building and satellite link terminal was recommended. Line suppressors and filters on all power, control, and data lines as they entered the control building or satellite terminal were to be added. The VLA facilities organization carried out the recommendations.

Lightning does strike twice. A videotape of the control building during a violent thunderstorm in August 1989 suggests that the newly installed lightning air terminal system intercepted a direct lightning strike and protected the installation without a detectable power surge or damage to equipment.

## G. System Availability

While an interagency array availability requirement of 80 percent was imposed for the PCTA Uranus-encounter configuration, the favorable experience for that encounter led to an implicit design goal on the order of 90 percent for the entire VGTA for the Neptune encounter. By applying the 98-percent requirement at both the VLA and Goldstone to the added telemetry-processing equipment at both sites, by setting a similar goal for the VLA "common" system, and by working to a goal of 99 percent for 25 out of the 27 VLA antennas on-line, an overall design expectation of 92 percent was established, subject to potential additional autophasing losses during bad weather.<sup>18</sup>

<sup>15</sup> J. Dorman, *Very Large Array Power System Facilities, Voyager-Neptune Encounter—1989 Implementation Plan* (internal document), Jet Propulsion Laboratory, Pasadena, California, February 11, 1987.

<sup>16</sup> *JPL-VLA Equipment Transfer Agreement For the VLA Power System Facilities* (internal document), Jet Propulsion Laboratory, Pasadena, California, January 31, 1989.

<sup>17</sup> R. Perley and K. Bartos, "Assessment and Recommendations Concerning VLA Site Lightning/Grounding Protection," JPL IOM 3330-88-018 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 23, 1988.

<sup>18</sup> *VGTA System Requirements and Design*, Document 1220-2 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 15, 1987.

spurious emissions within the band or at subharmonics thereof. However, frequency coordination procedures can prevent RFI from fixed transmitters that share the 8400-MHz to 8450-MHz frequency allocation. JPL and NRAO frequency coordinators can also manage to control, or at least limit, RFI from spurious emissions in this band through knowledge of, and coordination with, operators of potential emitters. At the VLA, frequency coordination with the White Sands Missile Range and several Air Force bases eliminated potential RFI from fixed and airborne transmitters during Voyager operations.

Radio astronomy requires wideband receivers, so the NRAO designed the HEMT front ends with instantaneous bandwidth from 7.9 to 8.9 GHz, which means the gain and noise performance degrade slowly below and above this range. Consequently, very strong signals in the 7.5- to 9.5-GHz range could interfere with the Voyager signal through gain compression or intermodulation in the wideband VLA front end.

Coordination with the nearby Langmuir Laboratory for Atmospheric Research of the New Mexico Institute of Mining and Technology (NMIMT) and with the National Center for Atmospheric Research (NCAR) avoided a potentially serious RFI problem during the two months around Neptune encounter. During joint NMIMT and NCAR radar measurements of thunderstorms, NCAR flew its airborne 9.4-GHz radar at distances adequate to avoid damaging the VLA HEMT front ends. Both organizations ceased emitting during the hours of VLA preparation for and reception of Voyager telemetry. Earlier tests had confirmed that the NMIMT 9-GHz fixed radar, located at a 10,000-foot elevation and 43 km from the VLA center, would not damage the VLA HEMT front ends.

Much closer to home, JPL and NRAO eliminated spurious RFI emitters located at the VLA. Most notably, the exciter in the 6.3-GHz transmitter of the communications satellite link to Goldstone produced spurious emissions within several of the VLA radio astronomy frequency bands and at subharmonic frequencies.

From the beginning, the NRAO had expressed concerns about radio interference originating from an Earth station transmitter located in the vicinity of the 27-antenna array. The VLA frequency bands ranged from P-band through Ka-band, covering the frequency range of 0.3 GHz to 24.0 GHz; satellites operate in either C-band (5.925 to 6.425 GHz) or Ku-band (14.0 to 14.5 GHz).

Examination of the above bands showed the potential for direct interference in the Ka-band if the satellite oper-

ated in the lower portion of C-band, and in the Ku-band if the satellite operated in the upper portion thereof. Thus, the lower portion of C-band and upper portion of Ku-band were to be avoided in selecting the satellite transmitter frequencies. The actual frequencies used were in the C-band at 6.385 GHz (primary) and 6.345 GHz (backup).

This selection of frequencies presented no direct interference potentials for any of the VLA bands. However, the fourth harmonics of these did fall close enough to the Ka-band to cause the NRAO to impose effective isotropic radiated power (EIRP) limits on the level of maximum transmitter radiation in the band from 24.50 to 24.70 GHz. The limits imposed were -7 dBW maximum total power in any 500-MHz band, with the satellite antenna to be located at least 150 meters from the nearest of the VLA antennas. Through the use of a fourth-harmonic filter in the output of the high-power transmitter, the estimated EIRP was -9 dBW with a transmitted power level of 750 watts at C-band.

Although concerns about the direct interference had been resolved, the effects of exciter reference frequencies of 329.53125 and 321.03125 MHz remained questionable. Computations of possible interfering harmonic frequencies showed the potential for interference in several of the VLA bands.

Shortly following the transmitter installation, testing at the transmitter for spurious radiation levels showed that indeed there was potential for interference in both P-band (90 cm) and L-band (18 to 21 cm). Although some tests were conducted by normal VLA observation with favorable results, it was not feasible to scan the number of potential interfering frequencies with sufficient sensitivity in a reasonable period of time. The NRAO therefore conservatively imposed restrictions on the use of the transmitter "during all scheduled VLA observing at P-band and at L-band in spectral line mode and for long integration continuum observations."

Since this restriction imposed hardships on the ability to test and maintain the VGTA, further studies were conducted with the view of RFI shielding of the exciters. These studies concluded that additional shielding of at least 30 dB was required over the range of frequencies from 300 MHz to 1.4 GHz. Such an enclosure, with certified compliance to specification, was installed, yielding the confidence to remove the exciter operating restriction. This significantly enhanced the reliability and operating convenience of the link by permitting continuous operation during the summer months of weekly, and then daily, Voyager passes.

and the Reed-Solomon threshold proved to be within the 1/2 dB model.<sup>10</sup> The "better than theory" conclusion for all these results, including the ungapped baseline, cast some doubt upon the calibration and, hence, upon the validity of the results. This discrepancy of a fraction of a decibel was later resolved as a misinterpretation of the analytical model with respect to the Reed-Solomon coding overhead factor.

One result of this simulation activity was greater concern at JPL regarding the expected impact of the data gap upon VGTA performance. On one hand, some Voyager planners learned of the gap for the first time, and on the other, the VGTA Project discovered that, while all of Voyager's Neptune data rates were doubly encoded, one-third of the 21.6-kbps frame was without Reed-Solomon protection. This segment consisted of uncompressed playback data from Voyager's tape recorder and was a prime consideration in the encounter data-recovery strategy. It was somewhat reluctantly accepted that there would be some snow<sup>11</sup> in the playback images (which were interlaced with error-free real-time images), depending upon Goldstone SSNR during the gap. Evaluation of actual performance would await the VLA stand-alone recordings of 1987 and the Neptune Dual Processor Program (NDPP) tests of 1988, discussed in Section V. Table 2 summarizes the results of modeling and simulation.

#### D. Satellite Communications Link

The long VGTA baseline led to consideration of a satellite communications link for transporting the telemetry data in real time from the VLA to Goldstone. A terrestrial microwave link in Australia for the Uranus encounter had performed exceptionally well, with less than 0.1 dB data degradation attributable to the microwave link. Based on the PCTA link performance, it was expected that a "video" satellite link would perform as well as the terrestrial microwave link had in Australia.

To verify the performance of a video link, the RCA Earth station located at Goldstone was used to make round-trip measurements to and from a domestic satellite. This link was normally used for space shuttle video transmissions. Tests were conducted to determine the round-trip time delay, the square-wave frequency response, the amplitude response, and the noise spectral density of the passband from 0.1 MHz to 7.0 MHz.

<sup>10</sup> S. Howard, *The VLA-Goldstone Telemetry Array* (internal document), Jet Propulsion Laboratory, Pasadena, California, February 5, 1987.

<sup>11</sup> Bit error rates in the range of 0.001 to 1 percent.

Additional testing was conducted by sending simulated telemetry data from Goldstone Signal Processing Center (SPC 10), via a fiber-optic link, to DSS 16 (the 26-m antenna) for transmission to the satellite and return to SPC 10, then measuring data degradation in terms of bit-error-rate performance. Results showed degradation to be less than 0.1 dB,<sup>12</sup> which had also been experienced in Australia. With DSS 14 (a 70-m antenna) simulating the VLA, DSS 15 (a 34-m antenna) simulating DSS 14, and with the RCA satellite link at DSS 16, the Voyager 2 spacecraft was tracked to demonstrate system performance.

Based on these results, a VGTA system requirements document<sup>13</sup> was developed and a contract awarded for a fully redundant link to include both dedicated transponders and Earth stations at each site. While the contractor resolved initial start-up problems of reliability and operability, extensive testing was conducted to determine link performance. After adding special 4.5-MHz low-pass nine-pole elliptical filters at both ends of the link to optimize the frequency response, expected results were obtained, with degradation of less than 0.1 dB.<sup>14</sup>

Because of the exposure to the New Mexico elements, particularly the wind, the VLA transmitting antenna was securely mounted on a concrete pad to ensure stability. The remainder of the equipment was housed in a transportable trailer. At Goldstone, the receiving antenna was also securely mounted on a concrete pad. The remainder of the equipment was housed within the SPC 10 communications facility. Link operating frequencies and siting considerations at the VLA were analyzed for possible radio-frequency interference with VLA operations.

#### E. Radio-Frequency Interference

Radio-frequency interference (RFI) could impact the reception of Voyager telemetry at any receiving site at any time, and while not specifically identified as a design issue, RFI was closely related to the HEMT low-noise amplifier (LNA) use and to the presence of the satellite communications transmitter.

Interference can occur from authorized emissions in or adjacent to the deep space frequency band, and also from

<sup>12</sup> S. S. Kent, "RCA Satellite Voyager Simulated Data Test Results," JPL IOM 3393-87-51 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 31, 1987.

<sup>13</sup> *VGTA System Requirements, VLA-Goldstone DSN Intersite Communications*, Document 1220-5 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 1, 1987.

<sup>14</sup> S. S. Kent, "VGTA SATCOM Test Results of 10/19/88," JPL IOM 3393-88-118 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 26, 1988.

The X-band aperture efficiencies of the antennas equipped with the JPL-designed feedhorn [13] were measured via VLA single-dish observations of radio sources of known flux density. Antenna sensitivities were found to be 0.110 K/jansky, giving aperture efficiencies of  $0.62 \pm 0.03$  [14]. These results combined with the system temperature measurements yielded an estimated zenith value of 49.7 dB/K for the gain-to-system-temperature (G/T) ratio of an individual antenna. At a 30-degree elevation angle, which would be characteristic of Voyager for the greater part of each day's pass, the total value of G/T would be 62.7 dB/K for 27 antennas added in phase. This result includes approximately 1 dB of loss resulting from quantization and the data gaps (discussed below) at the VLA. By comparison, a DSN 64-m antenna would have a G/T of 58.0 dB/K at the same elevation, so the HEMT-equipped VLA was predicted to be "worth" nearly three 64-m DSN antennas.

## B. Phasing 27 VLA Antennas

A key point glossed over above is that the 27 VLA antennas had to be summed in phase with one another. Random relative phases of the antennas would have reduced the VLA gain by approximately 7 dB—far more than the margin for receiving Voyager telemetry. Instrument phases change slowly, and extra path delays in the system are calibrated and subtracted out in real time. However, the fluctuating troposphere causes independent phase variations along the paths from the spacecraft to each individual antenna. This process can cause the antennas to lose phase coherence with one another and must be corrected by means of observations of a point radio source—in this case, Voyager. A correction is made in near-real time by determining the antenna phase adjustments that maximize the cross-correlated amplitude for the point source on each baseline. To "phase" the VLA during Voyager observations, a minimum broadband signal-to-noise ratio (SNR) on the Voyager signal was necessary for each cross-correlation. Early tests [7] showed that such a procedure was viable, and extensive tests and simulation refined the system design [15].

The NRAO's replacement of the VLA computers in late 1987 was a key ingredient in the signal combining, as it allowed the phase-determination procedure to make use of information on all 351 baselines of the VLA simultaneously, rather than to use only the 26 baselines to a single reference antenna. The enhancements in SNR achieved by means of both the computer improvements and the HEMT amplifiers were critical to the combining process. They allowed shorter integration times on the Voyager spacecraft to determine the phase adjustments; shorter integration

times meant more rapid feedback in the error-correction process. This, in turn, meant that the dynamic troposphere could be tracked even in the severe summer thunderstorms characteristic of the VLA location. During poor weather, the SNR takes a double hit. Increased system temperatures during thunderstorms make it more difficult to get enough SNR to determine the phases; at the same time, the phases fluctuate more rapidly, so the shortest possible integration time is desirable. Use of the HEMT amplifiers, rather than FET devices, was instrumental in keeping the signal-combining procedure robust even in the worst weather experienced during the encounter period.

## C. VLA Data Gap

Received signals at the antennas are returned for processing to the central electronics area by means of a buried waveguide system. Monitoring and control of the antennas, including the front-end electronics, are accomplished by time sharing through the same path. The resulting data interruption of 1.6 msec every 5/96 of a second was of immediate concern during initial planning. Studies were undertaken as early as 1982 to determine whether Voyager's concatenated coding scheme could tolerate these data gaps [6]. The Voyager data rates ranged from 4.8 to 21.6 kilobits per second (kbps), with convolutional inner code of rate 1/2, length 7, and Reed-Solomon outer code. Findings were that while the error rate on the inner code could be as high as about 1.5 percent at high SNR for the VLA signal processed in a stand-alone mode, the error-correction capability of the outer code would yield error-free performance comparable to standard performance, with less than 1 dB penalty. The projected design would perform somewhat better, at a typical symbol SNR (SSNR) of 3 dB, with equal contributions being made by the VLA and the Goldstone complex in the combining process. In this case, the convolutional bit error rate (BER) would be on the order of 0.001 percent, with an error-free threshold penalty of about 1/2 dB. Follow-up analysis was reported in [16].

Given confidence that the concatenated coding would bridge the gap, the VGTA Project went forward, undertaking a hardware evaluation of the gap effect, beginning late in 1985 and continuing through most of 1986. This evaluation consisted of a simulation of the partially gapped data stream processed through the telemetry system at the DSN Compatibility Test Area (CTA 21), followed by Reed-Solomon decoding and evaluation by the Voyager Project ground data system. Convolutional decoding results confirmed the modeling within statistical tolerances,<sup>9</sup>

<sup>9</sup> S. S. Kent, "VGTA Gap Testing at CTA 21," JPL IOMs 3393-86-22, 23, 33 (internal documents), Jet Propulsion Laboratory, Pasadena, California, March 1986.

preparations and operation. As eventually implemented, the X-band became the VLA's most sensitive observing band; it has seen significant use in astronomy applications, including several experiments with the Goldstone Solar System Radar. Some further insight into the management aspects of the VGTA Project is provided in [9].

The on-site engineering, installation, test, and operation of the Parkes-Canberra array were managed, as they were for the Uranus encounter, by the staff at the CDSCC of the Australian Space Office of the Department of Industry, Technology and Commerce, in cooperation with the Parkes (CSIRO) staff.

### III. System Design

A high-level block diagram of the VLA-Goldstone system is shown in Fig. 1. The essential elements are

- (1) X-band (8.4-GHz) reception at the VLA and Goldstone
- (2) total spectrum combining of the 27 VLA signals
- (3) carrier demodulation to baseband at both sites
- (4) Earth-satellite data transmission to Goldstone
- (5) standard baseband arraying of two or three antennas at Goldstone
- (6) baseband combining of the two subarrays (VLA and GDSCC) at Goldstone
- (7) convolutional decoding and signal processing, and data transmission to JPL
- (8) symbol-stream recording at both sites to back up the real-time system
- (9) symbol-stream playback and combining at Goldstone

More detailed diagrams and descriptions are available.<sup>6</sup> The Parkes-Canberra system configuration was similar,<sup>7</sup> except that Parkes had a single aperture and the terrestrial microwave rather than the satellite link was used. A radio science open-loop recording system shared the DSN facilities and front-end system at Parkes.

In concept, both telemetry systems closely resembled the Uranus encounter configuration. Several sections of

this article describe in some detail the peculiarities of the VLA configuration, and the "Implementation" section outlines most of the changes in the telemetry-processing subsystems that were new for Neptune encounter. The following synopses describe the outcome of several VGTA design issues identified in the formal review of 1987.<sup>8</sup> Each topic represented either a potential show-stopper or a decision point in the implementation path.

#### A. VLA Front-End Design

Several of the design issues pertained to the front-end or radio-frequency performance of the VLA as adapted to Voyager needs. In particular, they concerned the low-noise amplifier, figure of merit (G/T), and the VLA signal-combining (autophasing) efficiency.

The baseline plan for the VLA 8.4-GHz systems utilized cooled FET receivers on each of the 28 (27 in use, plus one spare) VLA antennas. The first three receivers installed in 1984 and 1985 were FET systems as described in [10]. However, during this period, the technology of low-noise high-electron-mobility transistor (HEMT) amplifier systems was developing rapidly. These receivers offered the promise of a significant improvement in system temperature for the VLA antennas when the antennas operated in this band. A cooperative program among the NRAO, JPL, General Electric Co., and Cornell University had begun in 1984 to develop HEMT devices for several applications. Tests of HEMT amplifiers on the VLA, beginning in 1986, confirmed an improvement of 30 percent in overall system noise temperature, corresponding to an increase of 1.5 dB in the expected overall sensitivity of the VLA [11]. Because of this substantial improvement, all VLA antennas were ultimately equipped with HEMT amplifiers for the Neptune encounter.

The performance of the individual VLA antennas was investigated extensively from 1985 to 1989. This was a critical component of monthly tests at the VLA site. The performance of both FET- and HEMT-equipped antennas was tested via observations of natural radio sources; the results of these tests played a major role in the decision to utilize HEMT amplifiers. Tip-curve measurements made in late 1986 and early 1987 showed that average zenith system temperatures were 45 to 50 kelvins for the FET devices and 30 to 35 kelvins for the HEMT amplifiers. Antennas equipped with HEMTs later in the project showed even better performance, as lower-noise amplifiers became available [12].

<sup>6</sup> *VGTA System Requirements and Design*, Document 1220-2 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 15, 1987.

<sup>7</sup> *PCTA System Requirements and Design, Update for Neptune Encounter*, Document 1200-2A (internal document), Jet Propulsion Laboratory, Pasadena, California, July 15, 1988.

<sup>8</sup> *VLA-Goldstone Array Formal Review* (internal document), Jet Propulsion Laboratory, Pasadena, California, July 1987.

was just in the process of contracting with the CSIRO to use Parkes in support of the Giotto mission to Comet Halley. Fortunately, the Giotto configuration at X-band (8.4 GHz) was compatible with Voyager's, and arrangements were made to share equipment and spacecraft-tracking time from late 1985 through the January and March 1986 encounters with Uranus and Comet Halley, respectively.

The Parkes-CDSACC Telemetry Array (PCTA) was thereby able to collect Voyager telemetry data at Parkes, transmit it in real time to the Canberra Deep Space Communications Complex (CDSACC) over a new microwave link, and combine it with the local DSN signal for improved data quality. Special recording capability was provided at each site so that data could be recovered at a later time in the event of outages of the real-time system. Following the Uranus encounter, the microwave link remained in place for cooperative use in astronomy and Very Long Baseline Interferometry (VLBI). The addition of Parkes to the Australian DSN complex<sup>1</sup> yielded an increase of approximately 50 percent in the CDSACC's reception capability for Uranus encounter [2].<sup>2</sup>

Meanwhile, in 1982, the interagency array studies identified the Very Large Array (VLA), near the Continental Divide in New Mexico, as a means of meeting the Neptune encounter reception needs.<sup>3</sup> This premier radio astronomy facility, operated by the National Radio Astronomy Observatory (NRAO) and sponsored by the National Science Foundation (NSF), consists of 27 antennas, each 25 m in diameter [3]. Configured in a "Y" arrangement with a 20-km radius, the array functions primarily as an astronomical mapping instrument with resolution comparable to that of optical telescopes. As was the case for the Uranus encounter, augmentation of the DSN in any given longitude (i.e., in Australia or the United States) would improve the data return for the total encounter by permitting data playback at a high rate from the Voyager tape recorder when the spacecraft was in view of the most sensitive (i.e., the arrayed) complexes. The VLA signal would combine with that of the Goldstone DSN complex to form the VLA-GDSACC Telemetry Array (VGTA), more than doubling the Goldstone Deep Space Communications Complex (GDSCC) capability, consisting of a

70-m and one or two 34-m antennas. The PCTA would be reinstated to help maintain the global communications capability with Voyager [4].<sup>4</sup> In addition, for the Neptune encounter, the DSN was upgraded with a new 34-m antenna at the Madrid complex, and each of the DSN's 64-m antennas was enlarged to a 70-m diameter.

Early DSN experience in telemetry arraying including the early planetary encounters of the two Voyager spacecraft has been summarized [5]. Table 1 outlines the arraying occurrences of that period, as well as the configurations for the more recent Voyager encounters.

## II. Project Management

Early planning, studies, and tests in 1983, which were based upon initial visits to the VLA and on an exchange of letters between the directors of the NRAO and JPL in 1982, led to an effective project start in 1984 [6-8]. Even though 1984 was a year of heavy activity at JPL in preparation for the Uranus encounter, including the PCTA implementation, the first test of a single antenna at 8.4 GHz with a prototype front-end installation was accomplished in December. JPL supplied the feedhorn, and NRAO's Central Development Laboratory at Charlottesville, Virginia, provided the cooled field-effect transistor (FET) low-noise amplifier, with other front-end electronics supplied by the VLA Electronics Division. By early 1985, a Memorandum of Agreement between the NSF and NASA, as well as a Management Plan for a joint JPL-NRAO VGTA Project, had been signed.<sup>5</sup> The JPL TDA Engineering Office would have responsibility for the overall planning and management of the project, supported by implementation/preparation managers in JPL and the NRAO. An operations manager within JPL's TDA Mission Support Office and DSN Operations Office would be appointed.

The project was charged with the design, implementation, and operation necessary to deliver 40 spacecraft passes of arrayed support for the Voyager-Neptune encounter. At least 40 more monthly tests were conducted from late 1984 through early 1989. As compensation to the NRAO and the radio astronomy community, NASA agreed to fund permanent X-band (8.4-GHz) installations on all 28 VLA antennas (this included a spare), as well as to bear all direct costs attributable to the VLA's Voyager-related

<sup>1</sup> At the Canberra site, the 64-m antenna was locally combined with one 34-m antenna.

<sup>2</sup> *PCTA System Requirements and Design, Including Subsystem Requirements*, Document 1200-2 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 15, 1984.

<sup>3</sup> J. W. Layland, D. A. Bathker, D. W. Brown, R. C. Clauss, and C. E. Kohlhaase, *Potential VLA/DSN Arrayed Support to the Voyager at Neptune* (internal document), Jet Propulsion Laboratory, Pasadena, California, May 12, 1983.

<sup>4</sup> *PCTA System Requirements and Design, Update for Neptune Encounter*, Document 1200-2A (internal document), Jet Propulsion Laboratory, Pasadena, California, July 15, 1988.

<sup>5</sup> *Management Plan for the VLA-GDSACC Telemetry Array Project*, Document 1220-1 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 15, 1985.

N91-11982

305247

p-28

## Interagency Telemetry Arraying for Voyager-Neptune Encounter

D. W. Brown  
TDA Engineering Office

W. D. Brundage  
National Radio Astronomy Observatory

J. S. Ulvestad  
Tracking Systems and Applications Section

S. S. Kent  
Telecommunications Systems Section

K. P. Bartos  
Radio Frequency and Microwave Subsystems Section

*The reception capability of the Deep Space Network (DSN) has been improved over the years by increasing both the size and number of antennas at each complex to meet spacecraft-support requirements. However, even more aperture was required for the final planetary encounters of the Voyager 2 spacecraft. This need was met by arraying one radio astronomy observatory with the DSN complex in the United States and another with the complex in Australia. Following a review of augmentation for the Uranus encounter, both the preparation at the National Radio Astronomy (NRAO) Very Large Array (VLA) and the Neptune encounter results for the Parkes-Canberra and VLA-Goldstone arrays are presented.*

### I. Introduction and Background

In August 1989, 12 years after launch, the Voyager 2 spacecraft encountered the planet Neptune and its moons. Imaging of the planet and the surface of Triton, the largest of the moons, ranked high on the list of experiments to be conducted over the several months surrounding closest approach. However, the steadily weakening signal received from the spacecraft as it receded from Earth tended to slow the rate at which images could be received by the Deep Space Network (DSN) and hence to reduce the quality and quantity of imaging data. Over the years, encoding modifications aboard Voyager 2 had been accomplished and the

ground reception capability had been steadily improved to regain communications capacity.

One approach was temporary augmentation of the DSN's own antennas at the Deep Space Communications Complexes around the world with suitable radio astronomy facilities. Beginning in 1981, studies were undertaken to identify, and negotiations to prepare, suitable candidates for the upcoming encounters with Uranus and then distant Neptune [1]. The first to receive serious attention was the Parkes Radio Telescope in Australia, operated by the Commonwealth Scientific and Industrial Research Organization (CSIRO). The European Space Agency (ESA)

tracking of radio sources. In its present design, the reflector can follow a zero-degree declination source for about one hour ( $\pm 30$  minutes) around its meridian transit by moving the focal feedhorn along a track from west to east. Tracking time increases in proportion to the secant of the declination. (The OSURT has a parabolic-shaped fixed reflector; it can track radio sources only about half the time that the NRT can.) The mobile feed is implemented by mounting the feedhorn and low-noise amplifiers in a small building, or shed, that rides on the track. The building is called a "chariot."

To limit spherical aberration in the system, the center of curvature of the spherical reflector has been set at a distance of 100 m behind the plane reflector. Furthermore, the spherical mirror is only illuminated over 200 m in the east-west direction. Also, the four outermost elements in the movable reflector are displaced slightly backwards to increase the electric path for rays striking the edges of the spherical reflector. This correction is analogous to the Schmidt correction used in optical systems.

The surfaces of the reflecting elements are made of 1.2 cm of welded mesh fixed on stretched cables. The root-mean-square accuracy of the surface is  $\pm 0.5$  cm from the nominal surface. The movable reflector can be positioned within  $\pm 0.1$  minutes of arc from a fixed reference.

Table 1 gives various parameters for the telescope and some pertinent information on the presently available receivers and back ends. Figure 2 shows the current sensitivity of the telescope in circular polarization receive mode at 21 cm wavelength.

## II. Potential Use of the NRT for SETI Observations

The general problem of searching for signals originating from extraterrestrial intelligence involves choosing search directions, frequencies, modulations, and sensitivities. Detailed studies carried out by several science working groups in the United States [1,2] reached the consensus that (1) the microwave window (1–10 GHz) is a logical and promising spectral region for interstellar communications, (2) both directed (target) searches and wide-area sky searches should be used, and (3) high-sensitivity telescope systems equipped with high-resolution multimillion-channel spectrum analyzers should be employed in the searches. Heidmann [3] suggested that the Nançay radio telescope could be successfully used with the NASA megachannel spectrum analyzers currently under development. The use of the NRT for SETI was considered with these factors in mind.

The NRT is the third largest radio telescope (after Arecibo, in Puerto Rico, and just after Effelsberg, in Germany) in existence; its effective area is equivalent to a circular aperture of 94 m. Table 2 compares the sensitivity of the NRT with other radio telescopes that have been considered for use by NASA: Arecibo, Green Bank, Ohio State, and all of the Deep Space Network 70-m and 34-m antennas. The sensitivity is given relative to that of Arecibo.

The NRT can track radio sources over approximately 80 percent of the sky, extending from  $-38^\circ$  declination up to the north pole. The telescope is able to track radio sources for an hour or more, and to scan large areas of the sky.

The NRT is inherently a broadband design, with an upper frequency cutoff near 5 GHz, determined by the accuracy of the reflecting surfaces. The primary operational frequency range of the telescope, 1–3 GHz, is in the frequency range considered to be ideally suited for SETI searches. Figure 3 shows the spectral utility function for SETI, which represents the frequency for maximum utility for interstellar communications. This function is the reciprocal of the free-space and quantum-noise temperatures of the background, multiplied by the square root of the frequency, to account for the minimum allowable bandwidth (see Figure 1.1, page 3 in [2]). The normal range of operation of the NRT is at the very best part of the entire microwave spectrum. As a result, the NRT shows great potential for use in SETI observations.

The NASA SETI program consists of a target-search mode and a sky-survey mode. This bimodal strategy is intended to cover a wide range of possibilities. The target mode assumes that other solar systems are similar to ours, and that intelligent life may form on planets around solar-like stars. The target-search strategy is to apply the highest possible sensitivity to detect putative radio signals coming from the direction of the nearest solar-type stars. The NRT is well suited to carrying out a target search because it has access to a large number of solar-type stars and it can track sufficiently long to enhance sensitivity. It has been estimated that over 500 F, G, or K stars within 100 light-years from the Sun can be examined with the NRT. Special galactic and extragalactic objects can also be included in a target search program.

The utility of the NRT for targeted search work has already been demonstrated. A collaborative search program these last years carried out by two of the authors (F. Biraud and J. Tarter) has concentrated on solar-type stars that lie beyond the declination limits of the Arecibo

telescope. These observations have used the existing 1024-channel autocorrelator to achieve a spectral resolution of 50 Hz over a bandwidth of 3 MHz centered around the molecular hydrogen (HI) line and lines of the hydroxyl radical (OH). Real-time reduction of the signals has not been possible, but efficient software has been written for the off-line computing system to allow the data taken on a given star to be completely analyzed before that star rises again. This capability and the generally favorable environment at Nançay within the protected radio astronomy bands enhance the efficiency of the observations. Previous non-real-time targeted searches required the antenna to be directed "off-source" for significant amounts of time to guard against false identification of interfering signals. At Nançay the percentage of time spent off-source can be reduced.

The sky-search mode allows for the possibility that intelligent life may exist in directions that we are not able to identify at the present time. The strategy is to search for radio signals over large areas of the sky. The small beam size of the NRT makes this strategy difficult, but not impossible, to carry out. Used as a transit instrument, the NRT can map approximately a 1/3 degree strip of sky in 24 hours at 21 cm, and can map the entire sky at a single frequency in 270 days. This is prohibitively long for the NRT, which is primarily devoted to astronomical work. The time required to map the sky in a specified frequency range can be shortened by

- (1) raster scanning the telescope beam in the sky
- (2) employing a multifrequency feed system
- (3) enlarging the telescope beam by underilluminating the primary mirror

Raster scanning can be achieved by rapidly moving the chariot (which holds the feedhorn) along the hour-angle scan track and/or by moving the flat reflector in a stepped or continuous motion. Raster scanning can easily reduce the survey times by factors of from 2 to 3, and perhaps up to 10.

The spherical reflector of the NRT permits a multibeam feed to be used with different declinations. Its availability on the NRT should allow timesharing of a large spectrum analyzer with several different feeds to speed up a search. Using this approach, survey times could probably be decreased by factors of from 2 to 3.

Increasing the size of the telescope beam defeats the purpose of using the large collecting area of the NRT; therefore, it is not an attractive alternative. However,

there may be interesting combinations involving multiple beams and receivers and a single-spectrum analyzer that permit the NRT to perform a sky survey at sensitivities that exceed those contemplated by the NASA SETI program. These possibilities are especially attractive at frequencies around 1.0 GHz, and deserve further studies.

Note that the NRT is particularly well suited to carry out surveys limited in space or frequency for which high sensitivity is desired. Particularly interesting examples are high sensitivity searches in the galactic plane, and searches in the water-hole region (1.4–1.7 GHz). The high-sensitivity galactic plane searches could be associated with searches for interesting galactic objects such as fast pulsars.

Finally, the NRT is very well suited for mounting an off-axis focal feed and receiver system. Such an off-axis system could be used 24 hours a day without interference to ongoing astronomical programs. While the declination of the search would be restricted by the astronomical program in progress at the time, the choice of frequencies could be freely determined by the requirements of SETI. This approach may not be the best use of the large NASA spectrum analyzer, but it could be implemented with other, less expensive, automated signal processors. This "com-mensal," or parallel detection strategy, is very attractive and has already been implemented to some degree with the SERENDIP system at the National Radio Astronomy Observatory (NRAO) [4]. SERENDIP is more restricted than what is proposed here for the NRT, because that system must accept not only the direction on the sky but, in addition, the frequency that has been chosen by the principal telescope user. Nevertheless, the project has been judged to have scientific merit, and has received funding from NASA. If it can be demonstrated that the parallel search at Nançay could significantly increase the volume of multidimensional space that can be investigated (e.g., in some interesting parameters such as frequency coverage, sensitivity, or repetition period for pulsed signals) there is a good possibility of obtaining NASA funding.

### III. Potential Upgrades for the Nançay Radio Telescope

While the NRT itself has significant potential for SETI observations, a number of improvements are necessary in order to take full advantage of the system in carrying out an efficient SETI program. In particular, an efficient SETI search must operate with the highest sensitivity available, and with the largest multichannel spectrum analyzers available. NASA has designed and built prototype

spectrum analyzers for SETI with upwards of one million channels. These analyzers increase the search speed that is possible over existing analyzers by a factor of more than 10,000.

The discussion that follows focuses on a number of potential improvements to the NRT that will make it compatible with the new class of spectrum analyzers with up to ten million channels which will become available over the next five years. This discussion is not intended to be exhaustive. It is the result of several meetings held between the authors of this article and the engineering staff at the NRT. Additional work is necessary to define and prioritize the upgrades.

### A. Aperture Efficiency

The current peak aperture efficiency of the telescope, operating in circular polarization at 21 cm wavelength, is only 45 percent. This yields a sensitivity of 1.1 K per jansky ( $1 \text{ Jy} = 10^{-26} \text{ W}/(\text{m}^2\text{Hz})$ ). It should be possible to increase the efficiency to 55 percent or more. Some work will need to be done to determine where the losses are originating from. Potential sources of losses are the reflecting surfaces, the illumination pattern on the primary surface, or possibly losses in the feedhorn that show up as aperture efficiency losses because of the calibration procedures that are being used. Preliminary discussions indicated that the Schmidt correction was optimized for 9 cm wavelength, but could be changed for longer wavelengths, to provide a maximum in the aperture efficiency of between 18 and 21 cm.

Additional increases in aperture efficiency can also be accomplished if the optics of the telescope are transformed from prime focus to a system employing a carefully shaped secondary reflector. In addition to providing increased bandwidth, a Gregorian reflector might also improve aperture efficiency by altering the illumination pattern of the primary mirror. Increased bandwidth and aperture efficiency cannot be achieved independently, but some optimum can be found that offers improvement in both parameters.

### B. System-Noise Temperature

Current system-noise temperatures in the 18- to 21-cm band are between 40 and 50 K. It should be possible to lower these temperatures by 25 percent or more by using lower noise receivers while decreasing the ground spillover using well-designed feeds, and by lowering the apparent temperature of the ground using reflecting screens. Plans already exist to erect additional rows of screen mesh on the ground in front of the spherical reflector.

The parameters, system-noise temperature and aperture efficiency, are not independent, and they should be optimized in such a way as to give the maximum sensitivity for the telescope. This condition implies that the ratio of system temperature to aperture efficiency should be minimal.

### C. Broadband Focal-Plane Feeds

A focal-feed system that spans 1–3 GHz is needed for SETI observations. The feeds should have an instantaneous bandwidth of at least 40 MHz to take advantage of the NASA target-mode spectrometer, and 30 MHz to take advantage of the NASA sky-survey spectrometer. The current feed system is limited to narrow ranges ( $\sim 150 \text{ MHz}$ ) around 21, 18, and 9 cm wavelengths. A Gregorian feed mounted a few meters behind the present feedhorn might facilitate the construction of a wideband feed system.

### D. Broadband RF and IF Receivers

Broadband, low-noise receivers to cover the 1–3 GHz spectral range will be needed.

### E. Spectrometers

Large multichannel spectrometers will need to be used in conjunction with the NRT in order to make efficient searches. Spectrometers currently under development in the United States for the NASA SETI program have more than 10,000 times the number of channels used in conventional radio astronomy spectrometers, such as those with 1,000 channels currently in use at the NRT. The possibility of using such a NASA spectrometer at the NRT should be investigated.

### F. Chariot

Since SETI instrumentation would compete for space in a chariot, it is necessary to consider the ways in which a chariot can be used. There are a number of possibilities for SETI.

First, it seems reasonable that SETI could share a chariot with radio astronomy by providing its own receivers. A disadvantage to this approach is that installation time would probably be needed to switch back and forth between radio astronomy and SETI.

A second concept is for SETI to provide a chariot dedicated to SETI observations to be parked out of the way when not in use. This dedicated-chariot concept would alleviate most interruptions to the radio astronomy operation.

A third concept is to use a stationary chariot for SETI, located either at one end of the track or elsewhere (e.g., above the track). This concept is interesting because the telescope could be used on a noninterference basis, nearly full time; however, it would at the same time greatly restrict the types of SETI programs that could be pursued. In particular, the fixed-chariot concept would rely on other programs for determination of the flat-mirror position, and hence the declination of the scans. As a result, target-mode observations could not track radio sources and would be restricted to those sources which pass through the declination setting. Similarly, the sky-survey observation would be restricted in declination. A further drawback for the fixed-chariot concept is that the receivers would not be available for radio astronomy.

An existing chariot, formerly used by the Centre National d'Etudes des Télécommunications (CNET) for ionospheric research, could possibly be used for either the movable, dedicated system, or a fixed system.

#### **G. Telescope Tracking**

Probably no modifications will be necessary to the tracking or pointing system. However, if a fast declination scanning mode of SETI observations were contemplated, it would be necessary to evaluate the limits set by mechanical and electrical aspects of the system.

#### **H. Computer System**

The current computer system is adequate to drive the telescope for SETI observations; however, it is somewhat out of date and there are plans to require funding to replace it. Additional computers will be needed for on-line

data analysis. These will be provided as part of the NASA spectrometers, and will be connected to the on-line NRT computer via a standard serial port.

### **IV. Radio-Frequency Interference**

As at any other radio astronomical site, the radio-frequency interference (RFI) level is a matter of concern. It should be monitored in order to track its possible origins, and hard-copy records should be kept for use in negotiations with the corresponding transmitter operators. In this respect, it is interesting to note that the Nançay Observatory has already started such monitoring with a dedicated dish. This should be strongly supported.

As a matter of fact, SETI, with its sophisticated signal-recognition procedures for data analysis, has the potential to reject most RFI; one way is to use Doppler shifts which extraterrestrial sources would not share with most terrestrial or even orbital transmitters.

Unfortunately, an increase in RFI level is expected; if it does occur, more efficient ways to distinguish RFI will have to be contemplated. One powerful way is to differentiate between Doppler shifts created by nearby moving transmitters (e.g., airplanes and satellites) and extraterrestrial sources. In the case of the NRT, a second element is currently being considered for conventional radio astronomy work. It may be a single dish, at least 30–40 m in diameter, located about 1 km from the NRT itself. Its interest for SETI RFI problems is plain. The need to use an interferometer for SETI may prove so compelling that some funds will need to be invested to make this possible.

## References

- [1] P. Morrison, J. Billingham, and J. H. Wolfe, eds., *The Search for Extraterrestrial Intelligence: SETI*, NASA SP-419, National Aeronautics and Space Administration, Washington, D.C., 1987.
- [2] F. Drake, J. H. Wolfe, and C. L. Seeger, eds., *SETI Science Working Group Report*, NASA-TP-2244, National Aeronautics and Space Administration, Washington, D.C., January 1984.
- [3] J. Heidman, "Le MegaSETI, une étape majeure en bioastronomie," *Journal des astronomes français*, no. 29, p. 11, 1987 (English translation in *New Ideas in Astronomy*, F. Bertola, B. Madore, and J. W. Sulentic, eds., New York: Cambridge University Press, 1988).
- [4] S. Bowyer, D. Werthimer, and V. Lindsay, "The Berkeley Piggyback SETI Program—SERENDIP II—Search for Extraterrestrial Radio Emission from Nearby Developed Intelligent Populations," *Bioastronomy—The Next Steps*, proceedings of the 99th IAU colloquium, Balaton, Hungary, June 22–27, 1987, G. Marx, ed., pp. 363–369, 1988.

**Table 1. Nançay radio telescope features**

Physical parameters			
Plane reflector:	Width	40 m	
	Length	200 m (east-west)	
	Height of elevation axis above ground	21 m	
Spherical reflector:	Height	35 m	
	Length	300 m (east-west)	
	Radius of curvature	560 m	
Distance between reflectors	460 m		
Total weight	7000 tons		
At focal plane	100 m = 20°		
Latitude	47° 22' 48'' N		
Longitude	8 min 47.4 sec E		
Performance parameters			
Declination coverage	-38° to + 90°		
Hour-angle coverage	±35 min x secant (declination)		
Collecting area	7000 m <sup>2</sup> maximum (decreasing north of 35° declination)		
Aperture efficiency	45% at 21 cm; 20% at 9 cm; 10% at 6 cm		
Beam width	~20 x 4' at 20 cm		
Tracking speed	2 cm/sec nominal		
Pointing accuracy	±0.5'		
Sensitivity	1.1 K/Jy at 21 cm; 0.5 at 9 cm; 0.25 at 6 cm		
Instrumentation			
Receivers			
Wavelength, cm	Frequency, GHz	Bandwidth, MHz	System temperature, K
21	1.4	150	40
18	1.6	150	45
9	3.3	300	65
Back ends			
Autocorrelator <sup>a</sup> :			
Channels	1024 (4 x 256, 2 x 512, or 1024)		
Bandwidth	6.4 MHz maximum		
Resolution	Down to 50 Hz		
Digitization	7-level (2 bits + sign) with weighting table		
<sup>a</sup> Designed by F. Biraud and D. A. Cesarsky.			

**Table 2. Relative sensitivities of radio telescopes**

Radio telescope	Diameter, m	System noise temperature, K	Aperture efficiency	Sensitivity
Nançay (now)	94	40	0.45	0.131
Arecibo	213	30	0.50	1.000
Green Bank	93	70	0.47	0.077
Ohio State	53	200	0.45	0.008
DSN 70 m	70	21	0.50	0.129
DSN 34 m	34	21	0.48	0.035

Notes: Sensitivity equals diameter squared times efficiency, divided by temperature, relative to Arecibo.

ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH

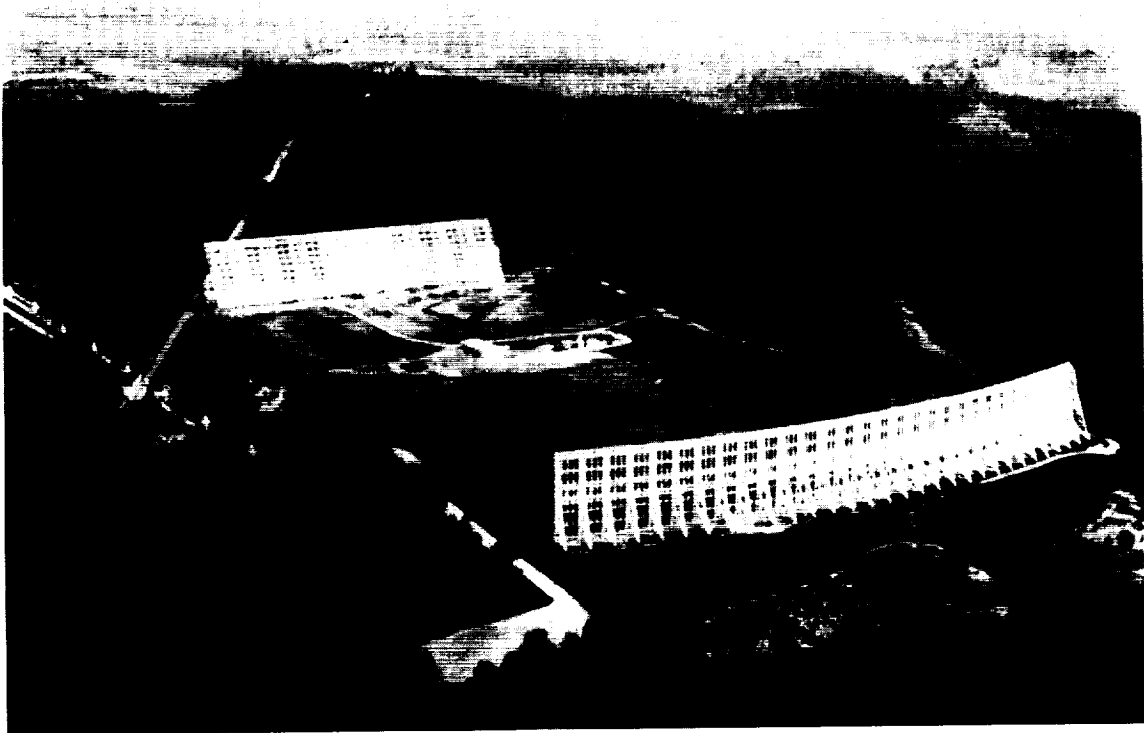


Fig. 1. Nancay Observatory.

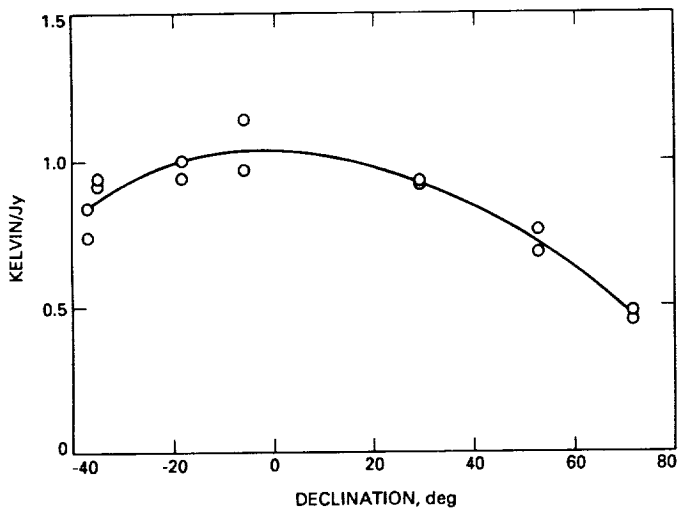


Fig. 2. NRT sensitivity in K/Jy versus declination in degrees.

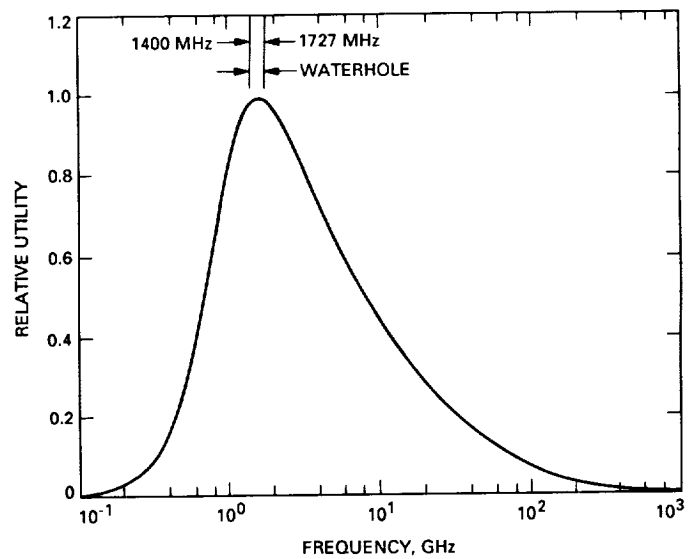


Fig. 3. Relative spectral utility function for SETI versus frequency in GHz.

## Referees

The following people have refereed articles for *The Telecommunications and Data Acquisition Progress Report*. By attesting to the technical and archival value of the articles, they have helped to maintain the excellence of this publication during the past year.

D. G. Bagby	C. Edwards	J. B. McNamee	E. H. Satorius
D. A. Bathker	P. Estabrook	A. Mileant	M. Shahshahani
D. Bayard	C. A. Greenhall	J. K. Miller	M. K. Simon
M. Calhoun	N. C. Ham	D. W. Murrow	L. Swanson
A. Cha	S. Hinedi	T. K. Peng	R. L. Sydnor
K.-M. Cheung	A. V. Kantak	S. M. Petty	S. W. Thurman
C. S. Christensen	P. Kinman	F. Pollara	R. N. Treuhaf
B. Clauss	R. Levy	E. C. Posner	J. S. Ulvestad
T. Cwik	S. M. Lichten	L. E. Primas	S.-C. Wu
L. J. Deutsch	K. M. Liewer	M. D. Rayman	L. E. Young
G. J. Dick	V. B. Lobb	N. A. Renzetti	
S. J. Dolinar	L. Maleki	G. M. Resch	
J. Dorman	M. McKenzie	R. Sadr	

